

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



(43) Date de la publication internationale
19 août 2004 (19.08.2004)

PCT

(10) Numéro de publication internationale
WO 2004/070705 A1

(51) Classification internationale des brevets⁷ :

G10L 19/02, 21/02

(21) Numéro de la demande internationale :

PCT/FR2004/000184

(22) Date de dépôt international :

27 janvier 2004 (27.01.2004)

(25) Langue de dépôt :

français

(26) Langue de publication :

français

(30) Données relatives à la priorité :

03/01081

30 janvier 2003 (30.01.2003) FR

(71) Déposant et

(72) Inventeur : CREBOUW, Jean-Luc [FR/FR]; 19 rue des
Bergères, Bât. Fructidor C, F-91940 Les Ulis (FR).

(74) Mandataire : DE SAINT PALAIS, Arnaud; Cabinet
Moutard, 35, rue de la Paroisse, F-78000 Versailles (FR).

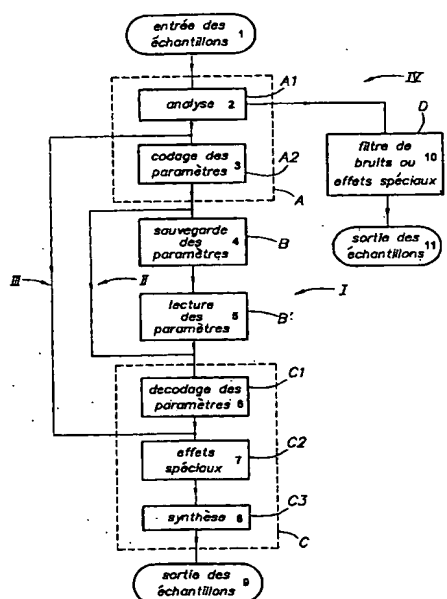
(81) États désignés (sauf indication contraire, pour tout titre de
protection nationale disponible) : AE, AG, AL, AM, AT,
AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO,
CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB,
GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG,
KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG,
MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH,
PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) États désignés (sauf indication contraire, pour tout titre de
protection régionale disponible) : ARIPO (BW, GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), eurasien
(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), européen (AT,

[Suite sur la page suivante]

(54) Title: METHOD FOR DIFFERENTIATED DIGITAL VOICE AND MUSIC PROCESSING, NOISE FILTERING, CREATION OF SPECIAL EFFECTS AND DEVICE FOR CARRYING OUT SAID METHOD

(54) Titre : PROCÉDE POUR LE TRAITEMENT NUMERIQUE DIFFERENCIÉ DE LA VOIX ET DE LA MUSIQUE, LE FILTRAGE DE BRUIT, LA CREATION D'EFFETS SPECIAUX ET DISPOSITIF POUR LA MISE EN OEUVRE DUDIT PROCÉDE



1 INPUT OF SAMPLES
2 ANALYSIS
3 ENCODING OF PARAMETERS
4 SAVING OF PARAMETERS
5 READING OF PARAMETERS
6 DECODING OF PARAMETERS
7 SPECIAL EFFECTS
8 SYNTHESIS
9 OUTPUT OF SAMPLES
10 NOISE FILTER OR SPECIAL EFFECTS
11 OUTPUT OF SAMPLES

(57) Abstract: The invention relates to a method for differentiated digital voice and music processing, noise filtering and the creation of special effects. The inventive method can be used to make the most of digital audio technologies, by performing a pre-encoding audio signal analysis, assuming that any sound signal during one frame interval is the sum of sines having a fixed amplitude and a frequency which is linearly modulated as a function of time, said sum being temporally modulated by the signal envelope and the noise being added to the signal prior to the sum.

(57) Abrégé : Le procédé pour le traitement numérique différencié de la voix et de la musique, le filtrage du bruit, et la création d'effets spéciaux selon l'invention, permet de tirer un meilleur parti des technologies audionumériques en effectuant préalablement au codage, une analyse du signal audio en considérant que tout signal sonore dans l'intervalle d'une trame est la somme de sinus d'amplitude fixe et dont la fréquence est modulée linéairement en fonction du temps, cette somme étant modulée temporellement par l'enveloppe du signal, le bruit étant rajouté à ce signal préalablement à ladite somme.



BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Publiée :

— avec rapport de recherche internationale

— avant l'expiration du délai prévu pour la modification des revendications, sera republiée si des modifications sont reçues

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

5 **PROCEDURE POUR LE TRAITEMENT NUMERIQUE DIFFERENCIE DE LA VOIX ET DE LA MUSIQUE, LE FILTRAGE DU BRUIT, LA CREATION D'EFFETS SPECIAUX ET DISPOSITIF POUR LA MISE EN ŒUVRE DUDIT PROCEDURE.**

10 La présente invention concerne un traitement numérique différencié de la voix et de la musique, le filtrage du bruit, la création d'effets spéciaux ainsi qu'un dispositif pour la mise en oeuvre dudit procédé.

15 Elle a plus particulièrement pour objet de transformer la voix d'une façon réaliste ou originale et, d'une manière plus générale, de traiter en temps réel la voix, la musique, le bruit ambiant et d'enregistrer les résultats obtenus sur un support informatique.

20 Elle s'applique notamment, mais non exclusivement, au grand public et aux professionnels du son qui souhaitent transformer la voix pour des applications ludiques, traiter différemment la voix et la musique, créer des effets spéciaux, réduire le bruit ambiant, et enregistrer, sous forme numérique compressée, les résultats obtenus.

25 D'une façon générale, on sait que le signal vocal est composé d'un mélange de signaux transitoires très complexes (bruits) et de parties de signal quasi-périodiques (sons harmoniques). Les bruits peuvent être de petites explosions : P, B, T, D, K, GU ; des bruits diffus doux : F, V, J, Z ou intenses CH, S ; quant aux sons harmoniques, leur spectre varie avec le type de voyelle et avec
30 le locuteur.

Les rapports d'intensité entre les bruits et les voyelles changent selon qu'il s'agit d'une voix conversationnelle, d'une voix parlée type conférence, d'une voix forte criée ou d'une voix chantée. La voix forte et la voix chantée favorisent les sons vocaliques au détriment des bruits.

5

Le signal vocal transmet simultanément deux types de messages : un message sémantique convoyé par la parole, expression verbale de la pensée, et un message esthétique perceptible au travers des qualités esthétiques de la voix (timbre, intonation, débit, etc.).

10

Le contenu sémantique de la parole, support d'une bonne intelligibilité est pratiquement indépendant des qualités de la voix ; il est véhiculé par les formes acoustiques temporelles ; une voix chuchotée n'est constituée que de bruits d'écoulement ; une voix « intime » ou de proximité est constituée d'un mélange de sons harmoniques dans les fréquences graves et de bruits d'écoulement dans l'aigu ; la voix d'un conférencier ou d'un chanteur possède un spectre vocalique harmonique riche et intense.

15

Quant aux instruments de musique, ils sont caractérisés par leur tessiture, c'est-à-dire l'étendue en fréquence de toutes les notes qu'ils peuvent émettre. Néanmoins très peu d'instruments possèdent un « son harmonique », c'est-à-dire un fondamental intense accompagné d'harmoniques dont l'intensité décroît avec le rang.

20

D'autre part, la tessiture musicale et le contenu spectral ne sont pas directement liés ; certains instruments ont les maxima d'énergie inclus dans la tessiture ; d'autres présentent une zone maximale d'énergie bien circonscrite, située à la limite aiguë de la tessiture et au-delà ; d'autres enfin ont des maxima d'énergie très étalés qui débordent largement la limite aiguë de la tessiture.

30

Par ailleurs, on sait que le traitement analogique de ces signaux complexes, par exemple leur amplification, provoque, d'une façon incontournable, une dégradation croissante au fur et à mesure dudit traitement et cela d'une manière irréversible.

5

L'originalité des technologies numériques est d'introduire un déterminisme (c'est-à-dire une connaissance a priori) le plus important possible au niveau des signaux traités de manière à réaliser des traitements particuliers qui résideront sous forme de calculs.

10

Ainsi, si l'on opère une transformation du signal représentatif d'un son, à l'origine sous sa forme naturelle de vibrations, en un signal numérique muni des propriétés évoquées précédemment, ce signal sera traité sans subir de dégradation telles que bruit de fond, distorsion et limitation de bande
15 passante ; de plus, il pourra être traité afin de créer des effets spéciaux tels que la transformation de la voix, la suppression du bruit ambiant, la modification du souffle de la voix, la différenciation de la voix et de la musique.

Bien entendu, la technologie audionumérique comprend les trois étapes
20 principales :

- la conversion du signal analogique en un signal numérique,
- les traitements souhaités transposés en équations à résoudre,
- la conversion du signal numérique en signal analogique puisque le dernier maillon de la chaîne génère des vibrations acoustiques.

25

D'une manière générale, on sait que les dispositifs de traitement du son, désignés sous le terme de vocodeur, comportent les quatre fonctions suivantes :

- l'analyse,
- 30 • le codeur,
- le décodeur,

- la synthèse.

Par ailleurs, les procédés de réduction de débit sont employés essentiellement pour le stockage numérique (dans le but de diminuer le volume binaire) et pour la transmission (dans le but de diminuer le débit nécessaire). Ces
5 procédés comprennent un traitement préalable au stockage ou à la transmission (codage) et un traitement à la restitution (décodage).

Parmi les procédés de réduction de débit, ceux utilisant les méthodes perceptuelles avec pertes d'information sont les plus employés et notamment
10 le procédé MPEG Audio.

Ce procédé repose sur l'effet de masque de l'audition humaine, c'est-à-dire la disparition des sons faibles en présence des sons forts, équivalent à un
15 déplacement du seuil d'audition provoqué par le son le plus fort et fonction de l'écart de fréquence et de niveau entre les deux sons.

Ainsi, le nombre de bits par échantillon est défini en fonction de l'effet de masque étant donné que les sons faibles et le bruit de quantification sont
20 inaudibles. Afin de tirer le meilleur profit de cet effet de masque, le spectre audio est divisé en un certain nombre de sous-bandes, permettant ainsi de préciser le niveau de masque dans chacune des sous-bandes et de réaliser une allocation binaire pour chacune d'entre elles.

25 Le procédé MPEG audio consiste ainsi à :

- numériser en 16 bits avec un échantillonnage à 48 kHz,
- déduire la courbe de masque entre 20 Hz et 20 kHz,
- diviser le signal en 32 sous bandes,
- évaluer le niveau maximal atteint dans chaque sous bande et durant
30 24 ms,
- évaluer le niveau de bruit de quantification juste inaudible,

- allouer le nombre de bits pour le codage,
- générer le nombre de bits dans la sous bande,
- encapsuler ces données en une trame de données qui se répète chaque 24 ms.

5

Cette technique consiste à transmettre un débit binaire variable selon la composition instantanée du son.

10 Néanmoins, ce procédé est plutôt adapté au traitement de la musique et non au signal vocal ; il ne permet pas de détecter la présence de la voix ou de la musique, de séparer le signal vocal ou musical et le bruit, de modifier la voix en temps réel pour synthétiser une voix différente mais réaliste, de synthétiser du souffle (bruit) pour créer des effets spéciaux, de coder un signal vocal comportant une seule voix, de réduire le bruit ambiant.

15

L'invention a donc plus particulièrement pour but de supprimer ces inconvénients.

20 Elle propose, à cet effet, un procédé permettant de tirer un meilleur parti des technologies audionumériques en effectuant préalablement au codage, une analyse du signal audio en considérant que tout signal sonore dans l'intervalle d'une trame est la somme de sinus d'amplitude fixe et dont la fréquence est modulée linéairement en fonction du temps, cette somme étant modulée temporellement par l'enveloppe du signal, le bruit étant rajouté à ce signal
25 préalablement à ladite somme.

Selon l'invention, ce procédé de transformation de la voix, de la musique et du bruit ambiant, fait essentiellement intervenir :

30 - durant la phase d'analyse :

- le calcul de l'enveloppe du signal,

- le calcul du « pitch » (période du fondamental du signal de voix) et de sa variation,
- l'application au signal temporel de la variation inverse du «pitch» par interpolation linéaire,
- 5 • la transformée rapide de Fourier (TRF) sur le signal prétraité,
- l'extraction des composantes fréquentielles et leurs amplitudes,
- le calcul du «pitch» et sa validation dans le domaine fréquentiel,
- l'élimination éventuelle du bruit ambiant par filtrage sélectif avant codage,

10

durant la phase de synthèse :

- la sommation des sinus dont l'amplitude des composantes fréquentielles varie en fonction de l'enveloppe du signal et dont les fréquences varient linéairement,
- 15 • le calcul des phases en fonction de la valeur des fréquences et des valeurs des phases et des fréquences appartenant à la trame précédente,
- la superposition du bruit,
- l'application de l'enveloppe.

20

Un mode d'exécution de l'invention sera décrit ci-après, à titre d'exemple non limitatif, avec référence aux dessins annexés, dans lesquels :

La figure 1 est un organigramme simplifié du procédé selon
25 l'invention ;

La figure 2 est un organigramme de l'étape d'analyse ;

La figure 3 est un organigramme de l'étape de synthèse ;

30

La figure 4 est un organigramme de l'étape de codage ; et

La figure 5 est un schéma synoptique d'un dispositif selon l'invention.

Dans cet exemple, le procédé de traitement numérique différencié de la voix et de la musique selon l'invention, représenté selon la figure 1, comprend les étapes suivantes :

- analyse du signal vocal (bloc A1),
- codage des paramètres (bloc A2),
- sauvegarde des paramètres (bloc B),
- 10 • lecture des paramètres (bloc B'),
- décodage des paramètres (bloc C1),
- effets spéciaux (bloc C2),
- synthèse (bloc C3).

15 Par ailleurs, l'analyse du signal vocal et le codage des paramètres constituent les deux fonctionnalités de l'analyseur (bloc A) ; de même, le décodage des paramètres, les effets spéciaux et la synthèse constituent les fonctionnalités du synthétiseur (bloc C).

Ces différentes fonctionnalités seront décrites ci-après, notamment en ce qui
20 concerne les différentes étapes constitutives des procédés d'analyse et de synthèse.

D'une manière générale, le procédé de traitement numérique différencié de la voix et de la musique comprend essentiellement quatre configurations de traitement :

- 25 • la première configuration (trajet I) comprenant l'analyse, suivie du codage des paramètres, suivi de la sauvegarde et de la lecture des paramètres, suivie du décodage des paramètres, suivi des effets spéciaux, suivis de la synthèse,
- la seconde configuration (trajet II) comprenant l'analyse, suivie du
30 codage des paramètres, suivi du décodage des paramètres, suivi des effets spéciaux, suivis de la synthèse,

- la troisième configuration (trajet III) comprenant l'analyse, suivie des effets spéciaux, suivis de la synthèse,
- la quatrième configuration (trajet IV) comprenant le filtre de bruit ou la génération d'effets spéciaux à partir de l'analyse, sans passer par la synthèse.

5

Ces différentes possibilités sont offertes à l'appréciation de l'utilisateur du dispositif mettant en œuvre le susdit procédé, lequel dispositif sera décrit ultérieurement.

10

Dans cet exemple, la phase d'analyse du signal audio (bloc A1), représentée selon la figure 2, comprend les étapes suivantes :

15

- mise en forme du signal d'entrée (bloc 1),
- calcul de l'enveloppe temporelle (bloc 2),
- détection d'interpolation temporelle (bloc 3),
- détection du signal audible (bloc 4),
- calcul de l'interpolation temporelle (bloc 5),
- calcul de la dynamique du signal (bloc 6),
- détection de trame inaudible après une trame d'énergie plus élevée (bloc 7),
- traitement d'impulsion (bloc 8),
- répétition de l'impulsion (bloc 9),
- calcul de la transformée rapide de Fourier (TRF) sur impulsion répétée (bloc 10),
- calcul des paramètres du signal servant au prétraitement avant la TRF (bloc 11),
- prétraitement du signal temporel (bloc 12),
- calcul de la TRF sur signal traité (bloc 13),
- calcul du rapport signal à bruit (bloc 14),
- test de la variation doppler du «pitch» (bloc 15),

20

25

30

- calcul de la TRF sur signal non traité (bloc 16),
- calcul du rapport signal à bruit (bloc 17),
- comparaison des rapports signal à bruit avec et sans prétraitement (bloc 18),
- 5 • restitution du résultat de la TRF avec prétraitement (bloc 19),
- calcul des fréquences et modules (amplitudes des composantes fréquentielles (bloc 20),
- décision du type de signal (bloc 21),
- test du 50 ou 60 Hz (bloc 22),
- 10 • calcul de la dynamique des modules dans le domaine fréquentiel (bloc 23),
- suppression de l'interpolation sur les données fréquentielles (bloc 24),
- suppression du signal inaudible (bloc 25),
- 15 • calcul et validation du «pitch» (bloc 26),
- décision si filtrage de bruit ou effets spéciaux, ou continuation de l'analyse (bloc 27),
- atténuation éventuelle du bruit ambiant (bloc 28),
- fin du traitement de la trame (bloc 29).

20

L'exploitation de la transformée rapide de Fourier (TRF) pour de la voix n'est pas envisageable étant donné la variabilité du signal fréquentiel ; en effet la variation des fréquences crée un étalement du résultat de ladite transformée rapide de Fourier (TRF) ; l'élimination de cet étalement est rendu possible

25 grâce au calcul de la variation du « pitch » et à l'application de la variation inverse dudit « pitch » sur le signal temporel.

Ainsi, l'analyse du signal vocal est effectuée essentiellement en quatre étapes :

- calcul de l'enveloppe du signal (bloc 2),
- 30 • calcul du «pitch» et de sa variation (bloc 12),

- application au signal temporel de la variation inverse du «pitch» (bloc 12),
- transformée rapide de Fourier (TRF) sur le signal prétraité (bloc 13),
- 5 • élimination éventuelle du bruit ambiant avant codage (blocs 23 à 28).

Par ailleurs, quatre seuils (blocs 4, 7, 8, 22) permettent de détecter respectivement la présence de signal inaudible, la présence de trame inaudible,
10 la présence d'une impulsion, la présence de signal perturbateur secteur (50 Hz ou 60 Hz).

D'autre part, un cinquième seuil (bloc 15) permet d'effectuer la transformée rapide de Fourier (TRF) sur le signal non traité en fonction des caractéristiques du «pitch» et de sa variation.

- 15 Un sixième seuil (bloc 18) permet de restituer le résultat de la transformée rapide de Fourier (TRF) avec prétraitement en fonction du rapport signal à bruit.

Enfin, une décision est prise (bloc 27) si le filtrage du bruit ou les effets spéciaux sont effectués ; dans le cas contraire, on continue l'analyse (flèche
20 IV).

Deux trames sont exploitées dans le procédé d'analyse du signal audio, une trame dite « courante », de périodicité fixe, contenant un certain nombre d'échantillons correspondant au signal vocal, et une trame dite « d'analyse »,
25 dont le nombre d'échantillons est équivalent à celui de la trame courante ou le double, et pouvant être décalée, en fonction de l'interpolation temporelle, par rapport à la susdite trame courante.

La mise en forme du signal d'entrée (bloc 1) consiste à effectuer un filtrage passe haut afin d'améliorer le codage futur des amplitudes fréquentielles en
30 augmentant leur dynamique ; ledit filtrage passe haut augmente la dynamique

d'amplitude fréquentielle en évitant qu'une fréquence basse audible n'occupe toute la dynamique et fasse disparaître des fréquences de faible amplitude mais néanmoins audibles. Le signal filtré est ensuite dirigé vers le bloc 2 pour la détermination de l'enveloppe temporelle.

5

Le calcul de l'enveloppe temporelle (bloc 2) permet de définir :

- le type de signal, s'il s'agit d'une impulsion avec ou sans signal de fond (bruit ambiant ou musique),
 - la position de la trame d'analyse de l'enveloppe du signal par rapport à la
- 10 trame courante,
- l'énergie du signal temporel.

Il est effectué par une recherche des maxima du signal, considérés comme la partie du «pitch» la plus élevée en valeur absolue.

- 15 On calcule ensuite le décalage temporel à appliquer à la trame d'analyse en recherchant d'une part le maximum de l'enveloppe dans ladite trame puis d'autre part deux indices correspondant aux valeurs de l'enveloppe inférieures d'un certain pourcentage à la valeur du maximum.

- Si dans une trame d'analyse on trouve localement un écart entre deux
- 20 échantillons supérieur à un pourcentage de la dynamique maximale de la trame et ce durant une durée limitée, on déclare qu'une impulsion brève est contenue dans la trame en forçant les indices de décalage temporels aux valeurs entourant l'impulsion additionnelle.

- 25 La détection d'interpolation temporelle (bloc 3) permet de corriger les deux indices de décalage de la trame d'analyse trouvés dans le calcul précédent, et ce en prenant en compte le passé.

- Un premier seuil (bloc 4) détecte ou non la présence d'un signal audible en
- 30 mesurant la valeur maximale de l'enveloppe ; dans l'affirmative, l'analyse de la trame est terminée ; dans le cas contraire, le traitement continu.

Un calcul est ensuite effectué (bloc 5) des paramètres associés au décalage temporel de la trame d'analyse en déterminant le paramètre d'interpolation des modules qui est égal au rapport de l'enveloppe maximale dans la trame courante à celle de la trame décalée.

La dynamique du signal est ensuite calculée (bloc 6) pour sa normalisation afin de diminuer le bruit de calcul ; le gain de normalisation du signal est calculé à partir de l'échantillon le plus élevé en valeur absolue dans la trame d'analyse.

Un second seuil (bloc 7) détecte ou non la présence d'une trame inaudible par effet de masque provoqué par les précédentes trames ; dans l'affirmative, l'analyse est terminée ; dans le cas contraire, le traitement continue.

Un troisième seuil (bloc 8) détecte ensuite ou non la présence d'une impulsion ; dans l'affirmative, un traitement spécifique est effectué (blocs 9, 10) ; dans le cas contraire, les calculs des paramètres du signal (bloc 11) servant au prétraitement du signal temporel (bloc 12) seront effectués.

En présence d'une impulsion, la répétition de l'impulsion (bloc 9) est effectuée en créant un « pitch » artificiel, égal à la durée de l'impulsion, de manière à éviter le masquage des fréquences utiles lors de la transformée rapide de Fourier (TRF).

La transformée rapide de Fourier (TRF) (bloc 10) est ensuite réalisée sur l'impulsion répétée en ne conservant que la valeur absolue du nombre complexe et non la phase ; le calcul des fréquences et des modules des données fréquentielles (bloc 20) est ensuite effectué.

En l'absence d'impulsion, le calcul des paramètres du signal (bloc 11) est effectué, lesquels paramètres concernent :

- le calcul du « pitch » et de sa variation,
- la définition du nombre d'échantillons dans la trame d'analyse.

En fait, le calcul du « pitch » est effectué préalablement par une différentiation
5 du signal de la trame d'analyse, suivi d'un filtrage passe bas des composantes
de rang élevé, puis d'une élévation au cube du résultat dudit filtrage ; la valeur
du « pitch » est déterminée par le calcul de la distance minimale entre une
portion de signal d'énergie élevée avec la suite du signal subséquent, étant
donné que la susdite distance minimale est la somme de la valeur absolue des
10 différences entre les échantillons du gabarit et les échantillons à corrélérer ;
ensuite, la partie principale d'un « pitch » centrée autour de une fois et demie
la valeur du « pitch » est recherchée en début de trame d'analyse afin de
calculer la distance de cette portion de « pitch » sur l'intégralité de la trame
d'analyse ; ainsi, les distances minimales définissent les positions des
15 « pitch », le « pitch » étant la moyenne des « pitches » détectés ; puis la
variation du « pitch » est calculée à l'aide d'une droite qui minimise l'erreur
quadratique moyenne des successions des « pitches » détectés ; le « pitch »
estimé en début et en fin de trame d'analyse en est déduit ; si le « pitch »
temporel en fin de trame est supérieur à celui en début de trame, la variation
20 du « pitch » est égale au rapport du « pitch » estimé de début de trame à celui
de fin de trame, diminué de 1 ; inversement, si le « pitch » temporel en fin de
trame est inférieur à celui en début de trame, la variation du « pitch » est égal à
1 diminué du rapport du « pitch » estimé en fin de trame à celui en début de
trame.

25

La variation du « pitch », trouvée et validée précédemment, sera soustraite du
signal temporel dans le bloc 12 de prétraitement temporel, en n'utilisant que le
premier ordre de ladite variation.

La soustraction de la variation du « pitch » consiste à échantillonner la trame d'analyse sur-échantillonnée avec un pas d'échantillonnage variant avec la valeur inverse de ladite variation du « pitch ».

5 Le sur-échantillonnage, dans un rapport deux, de la trame d'analyse est réalisé en multipliant le résultat de la transformée rapide de Fourier (TFR) de la trame d'analyse par le facteur $\exp(-j*2*PI*k/(2*L_trame))$, de manière à ajouter un délai d'un demi échantillon au signal temporel servant au calcul de la transformée rapide de Fourier ; la transformée rapide de Fourier inverse est ensuite réalisée afin d'obtenir le signal temporel décalé d'un demi échantillon.

10 Une trame de longueur double est ainsi réalisée en utilisant alternativement un échantillon de la trame originale avec un échantillon de la trame décalée d'un demi échantillon.

15 Après élimination de la variation du « pitch », ledit « pitch » semble identique sur toute la fenêtre d'analyse, ce qui donnera un résultat de la transformée rapide de Fourier (TRF) sans étalement de fréquences ; la transformée rapide de Fourier (TRF) pourra être ensuite effectuée dans le bloc 13 afin de connaître le domaine fréquentiel de la trame d'analyse ; la méthode utilisée
20 permet de calculer rapidement le module du nombre complexe au détriment de la phase du signal.

Le calcul du rapport du signal à bruit est effectué sur la valeur absolue du résultat de la transformée rapide de Fourier (TRF) ; le susdit rapport est en
25 fait le rapport de la différence de l'énergie du signal et du bruit à la somme de l'énergie du signal et du bruit ; le numérateur du susdit rapport correspond au logarithme de l'écart entre deux pics d'énergie, respectivement du signal et du bruit, le pic d'énergie étant celui qui est soit supérieur aux quatre échantillons adjacents correspondant au signal harmonique, ou inférieur aux quatre
30 échantillons adjacents correspondant au bruit ; le dénominateur est la somme des logarithmes de tous les pics du signal et du bruit ; par ailleurs, le calcul du

rapport signal à bruit se fait par sous-bande, les sous-bandes les plus élevées, en terme de niveau, sont moyennées et donnent le rapport recherché.

5 Le calcul du rapport de signal à bruit, défini comme étant le rapport signal moins le bruit à signal plus le bruit, effectué dans le bloc 14, permet de déterminer si le signal analysé est un signal voisé ou de la musique, cas d'un rapport élevé, ou du bruit, cas d'un rapport faible.

10 Cette distinction est ensuite effectuée dans le bloc 15 ; en fait, des tests sont effectués sur la variation doppler du « pitch » et sur la fréquence du « pitch » ; si la variation du « pitch » est faible ou sa fréquence élevée, le traitement est immédiatement suivi par le calcul des fréquences et des modules des données fréquentielles de la transformée rapide de Fourier (TRF) (bloc 20) ; dans le cas contraire, la transformée rapide de Fourier (TRF) est effectuée sans
15 prétraitement (bloc 16).

Le calcul du rapport signal à bruit est ensuite effectué dans le bloc 17, de manière à transmettre au bloc 20 les résultats de la transformée rapide de Fourier (TRF) sans prétraitement, cas d'une variation du « pitch » nulle, ou,
20 dans le cas contraire à restituer les résultats de la transformée rapide de Fourier (TRF) avec prétraitement (bloc 19).

Cette distinction est effectuée dans le bloc 18, de la manière suivante :

- si le rapport signal à bruit sans prétraitement est supérieur au rapport signal
25 à bruit avec prétraitement, les résultats de la transformée rapide de Fourier (TRF) sont transférés au bloc 20,
- si le rapport signal à bruit sans prétraitement est inférieur au rapport signal à bruit avec traitement, la restitution des résultats de la transformée rapide de Fourier (TRF) avec prétraitement étant effectuée dans le bloc 19, les
30 résultats obtenus avec prétraitement sont ensuite transférés dans le bloc 20.

Ce test permet de valider la variation du « pitch », qui pourrait être non nulle pour de la musique, alors que celle-ci doit effectivement être nulle.

- Le calcul des fréquences et des modules des données fréquentielles de la transformée rapide de Fourier (TRF) est effectué dans le bloc 20.
- La transformée rapide de Fourier (TRF), précédemment citée en référence aux blocs 10, 13, 16, est réalisée, à titre d'exemple, sur 256 échantillons dans le cas d'une trame décalée ou d'une impulsion, ou sur le double d'échantillons dans le cas d'une trame centrée sans impulsion.
- 10 Une pondération des échantillons situés aux extrémités des prélèvements, dite de HAMMING, est effectué dans le cas de la transformée rapide de Fourier (TRF) sur n échantillons ; sur $2n$ échantillons, on utilise la fenêtre de pondération de HAMMING multipliée par la racine carrée de la fenêtre de HAMMING.
- 15 A partir des valeurs absolues des données complexes de la transformée rapide de Fourier (TRF), on calcule le rapport entre deux valeurs maximales adjacentes, chacune représentant le produit de l'amplitude de la composante fréquentielle par un sinus cardinal ; par approximations successives, on compare ce rapport entre les valeurs maximales, à des valeurs contenues dans
- 20 des tableaux contenant ce même rapport, pour N fréquences (par exemple 32 ou 64) réparties uniformément sur un demi échantillon de la transformée rapide de Fourier (TRF). L'indice dudit tableau qui définit le rapport le plus proche de celui à comparer donne d'une part le module et d'autre part la fréquence pour chaque maximum de la valeur absolue de la transformée rapide
- 25 de Fourier (TRF).

Par ailleurs, le calcul des fréquences et des modules des données fréquentielles de la transformée rapide de Fourier (TRF), effectué dans le bloc 20, permet également de détecter un signal DTMF (multifréquence à double tonalité) en

30 téléphonie.

Il est à noter que le rapport signal à bruit est le critère essentiel qui définit le type de signal.

Afin de déterminer l'énergie du bruit à générer à la synthèse et la précision du codage, le signal extrait du bloc 20 est catégorisé en quatre types dans le bloc 21, à savoir :

- type 0 : signal voisé ou musique.

Le « pitch » et sa variation peuvent être non nuls ; le bruit appliqué à la synthèse sera de faible énergie ; le codage des paramètres sera effectué avec la précision maximale.

- type 1 : signal non voisé et éventuellement de la musique.

Le « pitch » et sa variation sont nuls ; le bruit appliqué à la synthèse sera de forte énergie ; le codage des paramètres sera effectué avec la précision minimale.

- type 2 : signal voisé ou musique.

Le « pitch » et sa variation sont nuls ; le bruit appliqué à la synthèse sera de moyenne énergie ; le codage des paramètres sera effectué avec une précision intermédiaire.

- type 3 : ce type de signal est décidé en fin d'analyse lorsque le signal à synthétiser est nul.

Une détection de présence ou de non présence de signal perturbateur à 50 Hz (60 Hz) est effectuée dans le bloc 22 ; le niveau du seuil de détection est fonction du niveau du signal recherché de manière à éviter de confondre la perturbation électromagnétique (50, 60 Hz) et la fondamentale d'un instrument de musique.

En présence du signal perturbateur recherché, l'analyse est terminée afin de diminuer le débit binaire : fin du traitement de la trame référencée par le bloc 29.

Dans le cas contraire, en l'absence de signal perturbateur, on continue l'analyse.

Un calcul de la dynamique des amplitudes des composantes fréquentielles, ou modules, est effectué dans le bloc 23 ; la susdite dynamique fréquentielle est utilisée pour le codage ainsi que pour la suppression des signaux inaudibles effectuée ultérieurement dans le bloc 25.

Ainsi, le plan fréquentiel est subdivisé en plusieurs parties, chacune d'entre-elles possède plusieurs plages d'amplitude différenciées en fonction du type de signal détecté au niveau du bloc 21.

D'autre part, l'interpolation temporelle et l'interpolation fréquentielle sont supprimées au niveau du bloc 24 ; celles-ci avaient été effectuées pour optimiser la qualité du signal.

L'interpolation temporelle qui donne des modules plus élevés, sera retirée en multipliant chaque module par le paramètre de normalisation calculé au niveau du bloc 5.

L'interpolation fréquentielle dépend de la variation du « pitch » ; celle-ci sera supprimée en fonction du décalage d'un certain nombre d'échantillons et du sens de la variation du « pitch ».

La suppression du signal inaudible est effectuée ensuite dans le bloc 25. En effet, certaines fréquences sont inaudibles car masquées par d'autres signaux d'amplitude plus élevées.

L'élimination de ces dites fréquences inaudibles permettra de diminuer le débit et aussi d'améliorer le calcul du « pitch » grâce à la suppression du bruit. Tout d'abord, on effectue une élimination des amplitudes situées en deçà de la limite inférieure de la plage d'amplitude, puis on éloigne les fréquences dont l'intervalle est inférieure à une unité fréquentielle, définie comme étant la fréquence d'échantillonnage par unité d'échantillon.

Ensuite, on élimine les composantes inaudibles à l'aide d'un test entre l'amplitude de la composante fréquentielle à tester et l'amplitude des autres

composantes adjacentes multipliée par un terme atténuateur fonction de la différence entre leur fréquence.

Par ailleurs, on limite le nombre de composantes fréquentielles à une valeur au-delà de laquelle la différence sur le résultat obtenu n'est pas perceptible.

5

Le calcul du « pitch » et la validation du « pitch » sont effectués au niveau du bloc 26 ; en effet le « pitch » calculé dans le bloc 11 sur le signal temporel a été déterminé dans le domaine temporel en présence de bruit ; le calcul du « pitch » dans le domaine fréquentiel permettra d'améliorer la précision du « pitch » et de détecter un « pitch » que le calcul sur le signal temporel, effectué dans le bloc 11, n'aurait pas déterminé à cause du bruit ambiant.

10

Par ailleurs, le calcul du « pitch » sur le signal fréquentiel doit permettre de décider si celui-ci doit être utilisé au codage, sachant que l'utilisation du « pitch » au codage permet de diminuer fortement le codage et de rendre la voix plus naturelle à la synthèse ; il est par ailleurs utilisé par le filtre de bruit.

15

Etant donné que les fréquences et les modules de la trame sont disponibles, le principe du calcul du « pitch » consiste à synthétiser le signal par une somme de cosinus ayant des phases à l'origine nulles ; ainsi la forme du signal original sera reconstitué sans les perturbations de l'enveloppe, des phases et de la variation du « pitch ».

20

La valeur du « pitch » fréquentiel est définie par la valeur du « pitch » temporel laquelle est équivalente à la première valeur de synthèse présentant un maximum supérieur au produit d'un coefficient par la somme des modules utilisés pour la synthèse locale (somme des cosinus desdits modules) ; ce coefficient est égal au rapport de l'énergie du signal, considéré comme harmonique, à la somme de l'énergie du bruit et de l'énergie du signal ; le susdit coefficient est d'autant plus faible que le « pitch » à détecter est noyé dans le bruit ; à titre d'exemple, à un rapport signal à bruit de 0 décibel correspond un coefficient de 0,5.

25

30

L'information de validation du « pitch » fréquentiel est obtenue à l'aide du rapport de l'échantillon de synthèse, à l'endroit du « pitch », à la somme des modules utilisés pour la synthèse locale ; ce rapport, synonyme d'énergie du signal harmonique sur l'énergie totale du signal, est corrigé en fonction du rapport approximatif signal à bruit calculé dans le bloc 14 ; l'information de validation du « pitch » dépend du dépassement du seuil de ce rapport.

Afin d'éviter de valider un « pitch » sur du bruit ou de la musique, quand le seuil de détection du « pitch » est faible, un contrôle de l'existence d'un « pitch » est effectué aux emplacements des multiples du « pitch » temporel dans la synthèse locale ; ainsi le « pitch » n'est pas valide si le niveau de la synthèse est trop faible pour être un « pitch » aux susdits emplacements des multiples du « pitch » temporel.

La synthèse locale est calculée deux fois ; une première fois en n'utilisant que les fréquences dont le module est élevé, afin de s'affranchir du bruit pour le calcul du « pitch » ; une deuxième fois avec la totalité des modules limités en valeur maximale, afin de calculer le rapport signal à bruit qui validera le « pitch » ; en effet la limitation des modules donne plus de poids aux fréquences non harmoniques à module faible, afin de diminuer la probabilité de validation d'un « pitch » sur de la musique.

20

Dans le cas du filtrage du bruit, les valeurs desdits modules ne sont pas limités pour la deuxième synthèse locale, seul le nombre de fréquences est limité en ne prenant en compte que celles qui ont un module significatif afin de limiter le bruit.

25

Un second procédé de calcul du « pitch » consiste à sélectionner le « pitch » qui donne l'énergie maximale pour un pas d'échantillonnage de la synthèse égal au « pitch » recherché ; ce procédé est utilisé pour de la musique ou un milieu sonore comportant plusieurs voix.

30

Préalablement à la dernière étape consistant à atténuer le bruit, une décision sera prise par l'utilisateur s'il souhaite réaliser le filtrage du bruit ou générer des effets spéciaux (bloc 27), à partir de l'analyse, sans passer par la synthèse.

Dans le cas contraire, l'analyse se terminera par le traitement suivant
5 consistant à atténuer le bruit, dans le bloc 28, en diminuant les composantes fréquentielles qui ne sont pas un multiple du « pitch » ; après atténuation desdites composantes fréquentielles, on effectuera à nouveau la suppression du signal inaudible, tel que décrit précédemment, au niveau du bloc 25.

L'atténuation desdites composantes fréquentielles est fonction du type de
10 signal tel que définit précédemment par le bloc 21.

Après avoir effectué ladite atténuation du bruit, on peut considérer que le traitement de la trame est terminé ; l'aboutissement de ladite phase d'analyse est référencé par le bloc 29.

15

En référence à la figure 1 représentant un organigramme simplifié du procédé selon l'invention, dans cet exemple, la phase de synthèse du signal audio (bloc C3), représentée selon la figure 3, comprend les étapes suivantes :

- mise en forme des modules (bloc 31),
- 20 • réduction du bruit (bloc 32),
- mise à niveau du signal (bloc 33),
- saturation des modules (bloc 34),
- modification des paramètres d'impulsion en fonction de la vitesse de la synthèse (bloc 35),
- 25 • calcul des phases (bloc 36),
- génération du souffle (bloc 37),
- décision concernant la génération d'une impulsion (bloc 38),
- synthèse avec les données fréquentielles de la trame courante (bloc 39),
- 30 • test concernant la trame précédente (bloc 40),

- synthèse avec les données fréquentielles de la trame précédente (bloc 41),
- application de l'enveloppe sur le signal de synthèse (bloc 42),
- décision concernant l'ajout d'une impulsion (bloc 43),
- 5 • synthèse avec les nouvelles données fréquentielles (bloc 44),
- connexion entre trames adjacentes (bloc 45),
- transfert du résultat de synthèse dans la trame d'échantillon (bloc 46),
- sauvegarde du bord de trame (bloc 47),
- 10 • fin de la synthèse (bloc 48).

La synthèse consiste à calculer les échantillons du signal audio à partir des paramètres calculés par l'analyse ; les phases et le bruit seront calculés artificiellement suivant le contexte.

15

La mise en forme des modules (bloc 31) consiste à éliminer l'atténuation du filtre d'entrée des échantillons de l'analyse (bloc 1 du bloc A1) et à tenir compte du sens de la variation du « pitch » car la synthèse est réalisée temporellement par un incrément de phase d'un sinus.

- 20 Par ailleurs, l'information de validation du « pitch » est supprimée si l'option de synthèse de la musique est validée ; cette option améliore le calcul de phase des fréquences en évitant de synchroniser les phases des harmoniques entre elles en fonction du « pitch ».

- 25 La réduction du bruit (bloc 32) est effectuée si celle-ci n'a pas été préalablement effectuée durant l'analyse (bloc 28 du bloc A1).

- La mise à niveau du signal (bloc 33) supprime la normalisation des modules reçus de l'analyse ; cette mise à niveau consiste à multiplier les modules par
- 30 l'inverse du gain de normalisation défini dans le calcul de la dynamique du signal (bloc 6 du bloc A1) et à multiplier lesdits modules par 4 afin d'éliminer

l'effet de la fenêtre de HAMMING, et que seule la moitié du plan fréquentiel est utilisée.

5 La saturation des modules (bloc 34) est effectuée si la somme des modules est supérieure à la dynamique du signal des échantillons de sortie ; elle consiste à multiplier les modules par le rapport de la valeur maximale de la somme des modules à la somme des modules, au cas où ledit rapport est inférieur à 1.

10 L'impulsion est re-générée en réalisant la somme de sinus dans la durée d'impulsion ; les paramètres d'impulsion sont modifiés (bloc 35) en fonction de la vitesse variable de synthèse.

15 Le calcul des phases des fréquences est effectué ensuite (bloc 36) ; il a pour but de donner une continuité de phase entre les fréquences des trames ou de re-synchroniser les phases entre elles ; elle rend par ailleurs la voix plus naturelle.

La synchronisation des phases est réalisée à chaque fois qu'un nouveau signal dans la trame courante semble séparé dans le domaine temporel ou dans le domaine fréquentiel de la trame précédente ; cette séparation correspond :

- 20
- au passage de signal bruité à un signal non bruité,
 - à un début de mot (ou son) dont l'enveloppe en début de trame est faible,
 - à une transition entre deux mots (ou son) sans variation de l'enveloppe,
- 25
- à un début de mot (ou son) qui a été détecté dans la trame précédente, mais dont la montée de l'enveloppe dans la trame courante est telle que la synchronisation doit être refaite pour que les phases soient calculées en fonction d'un « pitch » de meilleure qualité.

La continuité de phase consiste à rechercher les fréquences de la trame courante en début de trame qui sont les plus proches des fréquences en fin de trame de la trame précédente ; ensuite la phase de chaque fréquence devient égale à celle de la fréquence précédente la plus proche, sachant que les
5 fréquences en début de trame courante sont calculées à partir de la valeur centrale de la fréquence modifiée par la variation du « pitch ».

En présence d'un « pitch », cas du signal voisé, les phases des harmoniques seront synchronisées sur celle du pitch en multipliant la phase du « pitch » par
10 l'indice de l'harmonique du « pitch » ; quant à la continuité de phase, on calcule la phase du « pitch » en fin de trame en fonction de sa variation et de la phase à l'origine de la trame ; cette phase servira pour le début de la trame suivante.

15 Une seconde solution consiste à ne plus appliquer la variation du « pitch » sur le « pitch » pour connaître la nouvelle phase ; il suffit de reprendre la phase de la fin de la trame précédente du « pitch » ; par ailleurs, lors de la synthèse, la variation du « pitch » est appliquée sur l'interpolation de la synthèse réalisée sans variation du « pitch ».

20

La génération du souffle est ensuite effectuée (bloc 37).

Selon l'invention, on considère que tout signal sonore dans l'intervalle d'une trame est la somme de sinus d'amplitude fixe et dont la fréquence est modulée
25 linéairement en fonction du temps, cette somme étant modulée temporellement par l'enveloppe du signal, le bruit étant rajouté à ce signal préalablement à ladite somme.

Sans ce bruit, la voix est métallique car l'élimination des modules faibles, effectuée dans le bloc 25 du bloc A3, concerne essentiellement le souffle.

Par ailleurs, l'estimation du rapport signal à bruit effectuée dans le bloc 14 du bloc A3, n'est pas exploitée ; on calcule en effet un bruit en fonction du type de signal, des modules et des fréquences.

Le principe du calcul du bruit repose sur un filtrage d'un bruit blanc par un
5 filtre transversal dont les coefficients sont calculés par la somme des sinus des fréquences du signal dont les amplitudes sont atténuées en fonction des valeurs de leur fréquence et de leur amplitude. Une fenêtre de HAMMING est ensuite appliquée sur les coefficients pour diminuer les lobes secondaires.

Le bruit filtré est ensuite sauvegardé en deux parties distinctes.

10 Une première partie permettra de faire le lien entre deux trames successives ; la connexion entre deux trames est réalisée par chevauchement de ces deux trames dont chacune est pondérée linéairement et en sens inverse ; ledit chevauchement est effectué lorsque le signal est sinusoïdal ; il ne s'applique pas quand il s'agit de bruit non corrélé ; ainsi la partie sauvegardée du bruit
15 filtré est rajoutée sans pondération sur la zone de chevauchement.

La seconde partie est destinée au corps principal de la trame.

Le lien entre deux trames doit d'une part permettre un passage fluide entre deux filtres de bruit de deux trames successives, et d'autre part de prolonger le bruit de la trame suivante au-delà de la partie de chevauchement des trames si
20 un début de mot (ou son) est détecté.

Ainsi, le passage fluide entre deux trames est réalisé par la somme du bruit blanc filtré par le filtre de la trame précédente pondéré par une pente descendante linéaire, et le même bruit blanc filtré par le filtre de bruit de la trame courante pondéré par la pente montante inverse de celle du filtre de la
25 trame précédente.

L'énergie du bruit sera rajoutée à l'énergie de la somme des sinus, selon le procédé proposé.

La génération d'une impulsion diffère d'un signal sans impulsion ; en effet,
30 dans le cas de la génération d'une impulsion, la somme des sinus n'est réalisée

que sur une partie de la trame courante à laquelle est rajoutée la somme des sinus de la trame précédente.

- Cette distinction nécessite de faire le choix (bloc 38) entre les deux options :
- 5 une impulsion doit elle être générée ou non ? ; dans le cas où il n'y a pas de génération d'une impulsion , on effectue la synthèse avec les nouvelles données fréquentielles (bloc 39) ; dans le cas contraire, il s'agit de savoir si la trame précédente n'était pas une impulsion (bloc 40) ; dans ce cas on effectue la synthèse avec les données fréquentielles de la trame précédente (bloc 41)
- 10 qui va servir de fond à l'impulsion (cas de la musique ou de bruit ambiant à répéter) ; dans le cas contraire, la trame précédente étant une impulsion, on ne répète pas le signal de fond avec les paramètres de l'impulsion précédente.

- La synthèse avec les nouvelles données fréquentielles (bloc 39) consiste à
- 15 effectuer la somme des sinus des composantes fréquentielles de la trame courante ; la variation de la longueur de la trame permet d'effectuer une synthèse à vitesse variable ; néanmoins les valeurs des fréquences en début et en fin de trame doivent être identiques, quelque soit la longueur de la trame, pour une vitesse donnée de synthèse.

- 20 La phase associée au sinus, fonction de la fréquence, sera calculée par itération ; en effet pour chaque itération, on calcule le sinus multiplié par le module ; le résultat est ensuite sommé pour chaque échantillon suivant toutes les fréquences du signal.

- 25 Une autre méthode de synthèse consiste à réaliser l'inverse de l'analyse en recréant le domaine fréquentiel à partir du sinus cardinal réalisé avec le module, la fréquence et la phase, et ensuite en réalisant une transformée rapide de Fourier (TFR) inverse, suivie par le produit de l'inverse de la fenêtre de HAMMING pour obtenir le domaine temporel du signal.

- 30 Dans le cas où le « pitch » varie, l'inverse de l'analyse est à nouveau effectué en rajoutant la variation du « pitch » à la trame temporelle sur-échantillonnée.

Dans le cas d'une impulsion, il suffit d'appliquer au signal temporel, une fenêtre à 1 durant l'impulsion, et à 0, en dehors de celle-ci.

5 Dans le cas d'une impulsion à générer, les phases à l'origine des données fréquentielles sont maintenues à la valeur 0.

Afin de réaliser une connexion fluide entre les trames, le calcul de la somme des sinus est également effectué sur une portion précédant la trame et sur une même portion suivant la trame ; les parties aux deux bouts de la trame seront ensuite sommées avec celles des trames adjacentes par pondération linéaire.

10 Dans le cas d'une impulsion, la somme des sinus est effectuée dans l'intervalle de temps de génération de l'impulsion ; afin d'éviter la création d'impulsions parasites suite aux discontinuités dans le calcul de la somme des sinus, un certain nombre d'échantillons situés au début et à la fin de la séquence sont pondérés respectivement par une pente montante et une pente descendante.

15 Quant au cas des fréquences harmoniques du « pitch », les phases ont été calculées précédemment pour être synchronisées, elles seront générées à partir de l'indice de l'harmonique correspondant.

20 La synthèse par la somme des sinus avec les données de la trame précédente (bloc 41) est effectuée lorsque la trame courante contient une impulsion à générer ; en effet, dans le cas de la musique ou de bruit, si la synthèse n'est pas effectuée sur la trame précédente, servant de signal de fond, l'impulsion sera générée sur un silence, ce qui est préjudiciable à une bonne qualité du résultat obtenu ; par ailleurs la continuité de la trame précédente est inaudible,
25 même en présence d'une progression du signal.

L'application de l'enveloppe sur le signal de synthèse (bloc 42) est effectuée à partir des valeurs échantillonnées de l'enveloppe précédemment déterminées (bloc 2 du bloc A3) ; par ailleurs la connexion entre deux trames successives
30 est réalisée par la somme pondérée, comme indiqué précédemment ; cette

pondération par les courbes croissante et décroissante n'est pas effectuée sur le bruit, car le bruit n'est pas juxtaposé entre trame.

Enfin, dans le cas de la synthèse à vitesse variable, la longueur de la trame varie par pas afin d'être homogène avec l'échantillonnage de l'enveloppe.

5

Le rajout d'une impulsion par la somme de sinus dans l'intervalle où a été détectée l'impulsion, est effectué (bloc 44) en fonction du test réalisé précédemment (bloc 43).

10 La pondération de juxtaposition entre deux trames est ensuite effectuée (bloc 45) comme indiqué précédemment.

Le transfert du résultat de synthèse (bloc 46) sera ensuite effectué dans la trame de sortie d'échantillon afin que ledit résultat soit sauvegardé.

15

De même, la sauvegarde du bord de trame (bloc 47) sera effectuée afin que ledit bord de trame puisse être additionné au début de la trame suivante.

L'aboutissement de ladite phase de synthèse est référencée par le bloc 48.

20

En référence à la figure 1 représentant un organigramme simplifié du procédé selon l'invention, dans cet exemple, la phase de codage des paramètres (bloc A2), représentée selon la figure 4, comprend les étapes suivantes :

- codage du type de signal (bloc 51),
- 25 • test sur le type de signal (bloc 52),
- codage du type de compression (bloc 53),
- codage de la valeur de normalisation du signal de trame (bloc 54),
- test sur la présence d'impulsion (bloc 55),
- codage des paramètres d'impulsion (bloc 56),
- 30 • codage de la variation du « pitch » (bloc 57),
- limitation du nombre de fréquences à coder (bloc 58),

- codage des valeurs d'échantillonnage de l'enveloppe (bloc 59),
- codage de la validation du « pitch » (bloc 60),
- test de validation du « pitch » (bloc 61),
- codage des harmoniques (bloc 62),
- 5 • codage des fréquences non harmoniques (bloc 63),
- codage de la dynamique des modules (bloc 64),
- codage du module le plus élevé (bloc 65),
- codage des modules (bloc 66),
- codage de l'atténuation (bloc 67),
- 10 • suppression de la normalisation des modules (bloc 68),
- codage des fractions fréquentielles des fréquences non harmoniques (bloc 69),
- codage du nombre d'octets de codage (bloc 70),
- fin de codage (bloc 71).

15

Le codage des paramètres (bloc A2) calculés dans l'analyse (bloc A1) dans le procédé selon l'invention, consiste à limiter la quantité d'informations utiles afin de reproduire à la synthèse (bloc C3) après décodage (bloc C1) un équivalent auditif au signal audio d'origine.

- 20 Le codage étant de longueur variable, chaque trame codée a un nombre de bits d'information propre ; le signal audio étant variable, plus ou moins d'informations seront à coder.

Les paramètres de codage étant interdépendants, un paramètre codé influencera le type de codage des paramètres suivants.

25

Par ailleurs, le codage des paramètres peut être soit linéaire, le nombre de bits étant fonction du nombre de valeurs, soit de type HUFFMAN, le nombre de bits étant fonction statistique de la valeur à coder (plus la donnée est fréquente, moins elle utilise de bits et réciproquement).

30

Le type de signal, tel que défini lors de l'analyse (bloc 21 du bloc A1), fournit l'information de génération du bruit et la qualité du codage à utiliser ; le codage du type de signal est effectué en premier lieu (bloc 51).

- 5 Un test est ensuite effectué (bloc 52) permettant dans le cas du type 3 du signal, tel que défini dans le bloc 21 de l'analyse (bloc A1), de ne pas effectuer de codage des paramètres ; la synthèse comportera des échantillons nuls.
- 10 Le codage du type de compression (bloc 53) est utilisé dans le cas où l'utilisateur souhaite agir sur le débit des données de codage, au détriment de la qualité ; cette option peut être avantageuse en mode télécommunication associé à un taux de compression élevé.
- 15 Le codage de la valeur de normalisation (bloc 54) du signal de la trame d'analyse est de type HUFFMAN.

Un test sur la présence d'impulsion (bloc 55) est ensuite effectué, permettant en cas de synthèse d'une impulsion, de coder les paramètres de ladite
20 impulsion.

En cas de présence d'une impulsion, le codage, suivant une loi linéaire, des paramètres de ladite impulsion (bloc 56) sera effectué sur le début et la fin de ladite impulsion dans la trame courante.

25

Quant au codage de la variation doppler du « pitch » (bloc 57), il sera effectué suivant une loi logarithmique, en tenant compte du signe de la dite variation ; ce codage ne sera pas effectué en présence d'une impulsion ou si le type de signal est non voisé.

30

Une limitation du nombre de fréquences à coder (bloc 58) est ensuite effectuée afin d'éviter qu'une fréquence de valeur haute ne dépasse la dynamique bornée par la fréquence d'échantillonnage, étant donné que la variation doppler du « pitch » fait varier les fréquences durant la synthèse.

5

Le codage des valeurs d'échantillonnage de l'enveloppe (bloc 59) dépend de la variation du signal, du type de compression, du type de signal, de la valeur de normalisation et de l'éventuelle présence d'impulsion ; ledit codage consiste à coder les variations et la valeur minimale desdites valeurs d'échantillonnage.

10

La validation du « pitch » est ensuite codée (bloc 60), suivi d'un test de validation (bloc 61) nécessitant, dans l'affirmative, de coder les fréquences harmoniques (bloc 62) suivant leur indice par rapport à la fréquence du « pitch ». Quant aux fréquences non harmoniques, elles seront codées (bloc 63) suivant leur partie entière.

15

Le codage des fréquences harmoniques (bloc 62) consiste à effectuer un codage logarithmique du pitch, afin d'obtenir la même précision relative pour chaque fréquence harmonique ; le codage desdits indices des harmoniques est effectué en fonction de leur présence ou de leur absence par paquet de trois indices selon le codage d'HUFFMAN.

20

Les fréquences qui n'ont pas été détectées comme étant harmoniques de la fréquence du « pitch » seront codées séparément (bloc 63).

25

Afin d'éviter qu'au moment du codage, une fréquence non harmonique change de position par rapport à une fréquence harmonique, on supprime la fréquence non harmonique qui est trop proche de la fréquence harmonique, sachant qu'elle a moins de poids au sens audible ; ainsi la suppression a lieu si la fréquence non harmonique est supérieure à la fréquence harmonique et que la fraction de la fréquence non harmonique due au codage de la partie entière,

30

rend ladite fréquence non harmonique inférieure à la fréquence harmonique proche.

- Le codage des fréquences non harmoniques (bloc 63) consiste à coder le nombre de fréquences non harmoniques, puis la partie entière des fréquences, puis les parties fractionnaires quand les modules seront codés ; concernant le codage de la partie entière des fréquences, seuls les écarts entre lesdites parties entières sont codés ; par ailleurs, plus le module est faible et plus la précision sur la partie fractionnaire est faible ; ceci afin de diminuer le débit binaire. Afin d'optimiser le codage en terme de débit de la partie entière en fonction de la statistique des écarts de fréquence, on définit un certain nombre d'écarts maximal entre deux fréquences.

- Le codage de la dynamique des modules (bloc 64) utilise une loi de HUFFMAN en fonction du nombre de plages définissant ladite dynamique et du type de signal. Dans le cas d'un signal voisé, l'énergie du signal se situe dans les basses fréquences ; pour les autres types de signal, l'énergie est répartie uniformément dans le plan fréquentiel, avec une baisse vers les hautes fréquences.

- Le codage du module le plus élevé (bloc 65) consiste à coder, suivant une loi de HUFFMAN, la partie entière dudit module le plus élevé en tenant compte de la statistique dudit module le plus élevé.

- Le codage des modules (bloc 66) n'est réalisé que si le nombre de module à coder est supérieur à 1, étant donné que dans le cas contraire, il est seul en étant le module le plus élevé.

- Lors de l'analyse (bloc A1), la suppression du signal inaudible (bloc 25 du bloc A1) élimine les modules inférieurs au produit du module par l'atténuation correspondante ; ainsi un module se situe obligatoirement dans une zone du plan module/fréquence dépendant de la distance qui le sépare de ses deux modules adjacents en fonction de l'écart de fréquence desdits modules

adjacents. Ainsi la valeur du module est approximée relativement au module précédent en fonction de l'écart de fréquence et de l'atténuation correspondante qui dépend du type de signal, de la valeur de normalisation et du type de compression ; ladite approximation de la valeur du module est effectuée en référence à une échelle dont le pas varie suivant une loi logarithmique.

Le codage de l'atténuation (bloc 67) apportée par le filtre d'entrée des échantillons est effectué, puis est suivi de la suppression de la normalisation (bloc 68) qui permet de recalculer le module le plus élevé ainsi que la fréquence correspondante.

Le codage des fractions fréquentielles des fréquences non harmoniques (bloc 69) complète le codage des parties entières desdites fréquences.

La précision du codage va dépendre :

- de la fréquence : plus la fréquence est faible, plus la précision sera élevée de manière à ce que le rapport erreur de codage sur fréquence soit faible,
- du type de signal,
- du type de compression,
- de la valeur de normalisation du signal : plus l'intensité du signal est élevée, plus le codage est précis.

Enfin, le codage du nombre d'octets de codage (bloc 70) est effectué à l'issue du codage des différents paramètres susmentionnés, mémorisés dans une mémoire de codage dédiée.

L'aboutissement de ladite phase de codage est référencé par le bloc 71.

En référence à la figure 1 représentant un organigramme simplifié du procédé selon l'invention, dans cet exemple, la phase de décodage des paramètres est représentée par le bloc C1.

Le décodage étant l'inverse du codage, l'exploitation des bits de codage des
5 différents paramètres susmentionnés permettra de retrouver les valeurs d'origine des paramètres, avec d'éventuelles approximations.

En référence à la figure 1 représentant un organigramme simplifié du procédé selon l'invention, dans cet exemple, la phase de filtrage du bruit et de
10 génération d'effets spéciaux, à partir de l'analyse, sans passer par la synthèse est indiquée par le bloc D.

Le filtrage du bruit est effectué à partir des paramètres de la voix calculés dans l'analyse (bloc A1 du bloc A), empruntant le trajet IV indiqué sur ledit
15 organigramme simplifié du procédé selon l'invention.

Il s'avère que les algorithmes connus de l'état de l'art réalisent une annulation du bruit à partir des propriétés statistiques du signal ; le bruit doit être par conséquent stationnaire statistiquement ; cette démarche n'autorise donc pas la
20 présence de bruit sous forme harmonique (voix, musique).

L'objectif du filtrage du bruit est par conséquent de réduire toutes sortes de bruit tel que : bruit ambiant de voiture, de moteur, de foule, de musique, d'autres voix si celles-ci sont plus faibles que celles à conserver, ainsi que les
25 bruits de calcul de tout vocodeur (à titre d'exemple : ADPCM, GSM, G723).

Par ailleurs, la majorité des bruits ont leur énergie dans les basses fréquences ; le fait d'utiliser le signal de l'analyse préalablement filtré par le filtre d'entrée des échantillons permet de diminuer d'autant le bruit très basse fréquence.

Le filtrage du bruit (bloc D) pour un signal voisé consiste à réaliser la somme pour chaque échantillon, du signal original, du signal original décalé de un « pitch » en valeur positive et du signal original décalé de un « pitch » en valeur négative. Ceci nécessite de connaître pour chaque échantillon, la valeur du « pitch » et de sa variation. Avantageusement les deux signaux décalés sont multipliés par un même coefficient, et le signal original non décalé par un second coefficient ; la somme dudit premier coefficient rajouté à lui-même et dudit second coefficient est égale à 1, diminué de manière à conserver un niveau équivalent du signal résultant.

10

Le nombre d'échantillons espacés d'un « pitch » temporel ne se limite pas à trois échantillons ; plus il y a d'échantillons utilisés pour le filtre de bruit, et plus le filtre diminue le bruit.

Le nombre de trois échantillons est adapté au « pitch » temporel le plus élevé rencontré dans de la voix et au retard de filtrage. Afin de garder un retard de filtrage fixe, plus le « pitch » temporel est faible, plus on peut utiliser d'échantillons décalés d'un « pitch » pour réaliser le filtrage ; ce qui revient à garder la bande passante autour d'un harmonique, à peu près constante ; plus la fondamentale est élevée et plus la largeur de bande atténuée est élevée.

20

Par ailleurs, le filtrage du bruit ne concerne pas les signaux sous forme d'impulsion ; il est donc nécessaire de détecter la présence d'éventuelles impulsions dans le signal.

25 Le filtrage du bruit (bloc D) pour un signal non voisé consiste à atténuer ledit signal par un coefficient inférieure à 1.

Dans le domaine temporel, la somme des trois signaux sus mentionnés est corrélée ; quant au bruit contenu dans le signal original, la somme atténuera son niveau.

30

Ainsi, il est nécessaire de connaître avec exactitude la variation du « pitch », c'est-à-dire la valeur temporelle du « pitch », approximée en valeur linéaire, sachant qu'elle fait intervenir un terme du second ordre ; l'amélioration de la précision des deux susdits décalages, positif et négatif, est obtenue grâce à l'utilisation de la corrélation par la distance en début, milieu et fin de trame ; cette démarche a été décrite au cours de l'étape "calcul des paramètres du signal" (bloc 11 du bloc A1).

Avantageusement, le filtrage de bruit, décrit précédemment, permet de générer des effets spéciaux ; ladite génération d'effets spéciaux permet d'obtenir :

- une féminisation de la voix, en divisant la valeur temporelle du « pitch » par deux, pour certaines valeurs des amplitudes du signal original et des signaux originaux décalés ; ceci multiplie artificiellement la fréquence du « pitch » de la voix par deux en supprimant les harmoniques impaires ;
- une voix artificielle et étrange, en divisant la valeur temporelle du « pitch » par deux, pour d'autres valeurs des amplitudes du signal original et des signaux originaux décalés ; ceci permet de ne garder que les harmoniques impaires ;
- deux voix différentes, en divisant la valeur temporelle du « pitch » par deux, pour différentes valeurs des amplitudes du signal original et des signaux originaux décalés ; ceci permet d'atténuer les harmoniques impaires.

Enfin, une autre démarche, semblable à celle décrite précédemment permettant le filtrage du bruit, pourra être appliquée, non pas pour filtrer le bruit, mais pour diviser par deux ou par trois la fondamentale de la voix et ce, sans modification du formant (enveloppe spectrale) de ladite voix.

Le principe de ladite démarche consiste :

- à multiplier chaque échantillon de la voix originale par un cosinus variant au rythme de la moitié de la fondamentale (multiplication par deux du nombre de fréquences), ou variant au rythme du tiers de la fondamentale (multiplication par trois du nombre de fréquences),
- puis à additionner le résultat obtenu à la voix originale.

Par ailleurs, la phase de filtrage du bruit et de génération d'effets spéciaux, à partir de l'analyse, sans passer par la synthèse, peut ne pas inclure le calcul de la variation du « pitch » ; ceci permet d'obtenir une qualité auditive voisine de celle précédemment obtenue selon le procédé susmentionné ; dans ce mode opératoire, les fonctions définies par les blocs 11, 12, 15, 16, 17, 18, 19, 25 et 28 sont supprimées.

En référence à la figure 1 représentant un organigramme simplifié du procédé selon l'invention, dans cet exemple, la phase de génération d'effets spéciaux, associée à la synthèse (bloc C3) est indiquée par le bloc C2 du bloc C.

La dite phase de génération d'effets spéciaux, associée à la synthèse, permet de transformer la voix ou la musique :

- soit en modifiant selon certaines lois, les paramètres décodés issus du bloc C1 (trajet II),
- soit en traitant directement les résultats de l'analyse issus du bloc A1 (trajet III).

Les paramètres modifiés sont :

- le « pitch »,
- la variation du « pitch »,
- la validation du « pitch »,
- le nombre de composantes fréquentielles,
- les fréquences,
- les modules,

- les indices.

Les fréquences étant distinctes entre elles, leur transformation permet de rajeunir la voix, de la vieillir, de la féminiser ou inversement, de la transformer
5 en une voix artificielle. Ainsi la transformation des modules autorise toute sorte de filtrages et permet en outre de conserver la voix naturelle en gardant le formant (enveloppe spectrale).

A titre d'exemples, trois types de transformation de la voix seront décrits ci-
10 après, chacun étant référencé sous une appellation qui lui est propre, à savoir :

- la fonction "Transform" modifiant la voix de façon artificielle et permettant de créer un effet de chorale,
- la fonction "Transvoice" modifiant la voix de façon réaliste,
- la fonction "Formant" associée à la fonction "Transvoice".

15

La fonction "Transform" consiste à multiplier toutes les fréquences des composantes fréquentielles par un coefficient. Les modifications de la voix sont fonction de la valeur de ce coefficient, à savoir :

- une valeur supérieure à 1 transforme la voix en voix de canard,
- 20 • une valeur faiblement supérieure à 1 rajeunit la voix,
- une valeur inférieure à 1 rend la voix plus grave.

En effet, ce rendu artificiel de la voix est dû au fait que les modules des composantes fréquentielles sont inchangés et que l'enveloppe spectrale est
25 déformée.

Par ailleurs, en synthétisant plusieurs fois les mêmes paramètres, modifiés par ladite fonction "Transform" avec un coefficient différent, on réalise un effet de chorale en donnant l'impression que plusieurs voix sont présentes.

30 La fonction "Transvoice" consiste à recréer les modules des harmoniques à partir de l'enveloppe spectrale, les harmoniques originaux sont abandonnés

sachant que les fréquences non harmoniques ne sont pas modifiées ; à ce titre, ladite fonction "Transvoice" fait appel à la fonction "Formant" qui détermine le formant.

Ainsi, la transformation de la voix est effectuée de façon réaliste car le
5 formant est conservé ; un coefficient de multiplication des fréquences harmoniques supérieure à 1 rajeunit la voix, voire la féminise ; réciproquement, un coefficient de multiplication des fréquences harmoniques inférieure à 1 rend la voix plus grave.

Par ailleurs, afin de conserver un niveau sonore constant, indépendamment de
10 la valeur du coefficient de multiplication, les nouvelles amplitudes seront multipliées par le rapport de la somme des modules en entrée de ladite fonction "Transvoice" à la somme des modules en sortie.

La fonction "Formant" consiste à déterminer l'enveloppe spectrale du signal
15 fréquentiel ; elle est exploitée pour garder les modules des composantes fréquentielles constants quand les fréquences sont modifiées.

La détermination de l'enveloppe est effectuée en deux étapes, à savoir :

- un filtrage des modules placés dans l'enveloppe,
- une interpolation logarithmique de l'enveloppe entre deux modules
20 d'un harmonique.

Ladite fonction "Formant" peut être appliquée lors du codage des modules, des fréquences, des plages d'amplitudes et des fractions de fréquences, en n'effectuant le dit codage que sur les paramètres essentiels du formant, le
25 « pitch » étant validé. Dans ce cas, lors du décodage, les fréquences et les modules sont recalculés à partir respectivement du « pitch » et de l'enveloppe spectrale. Ainsi le débit binaire est réduit ; néanmoins, cette démarche n'est applicable qu'à la voix.

Les dites fonctions "Transform" et "Transvoice", décrites précédemment font intervenir un coefficient de multiplication des fréquences constant. Cette transformation peut être non linéaire et permettre de rendre la voix artificielle.

- 5 En effet, si ce coefficient de multiplication est fonction du rapport entre le nouveau « pitch » et le « pitch » réel, la voix sera caractérisée par un « pitch » fixe et un formant variable ; elle sera ainsi transformée en voix de robot associée à un effet spatial.
- 10 Si ce coefficient de multiplication varie périodiquement ou aléatoirement, à basse fréquence, la voix est vieillie associée à un effet hilarant.

Ces différentes transformations de la voix, obtenues à partir d'une modification, constante ou variable dans le temps, des fréquences, ladite
15 modification étant effectuée sur chacune des fréquences prises séparément, sont données à titre d'exemples.

Une dernière solution consiste à effectuer un codage à débit fixe. Le type de signal est ramené à du signal voisé (type 0 et 2 avec la validation du « pitch »
20 à 1), ou à du bruit (type 1 et 2 avec la validation du « pitch » à 0). Le type 2 étant pour la musique, il est éliminé dans ce cas, puisque ce codage ne peut coder que de la voix.

Le codage à débit fixe consiste à :

- 25 • coder le type de signal, l'information de la présence d'impulsion, et la validation du « pitch » en codage de HUFFMAN,
- coder l'emplacement de l'impulsion dans la trame si on n'est pas en présence d'une impulsion, sinon coder les parties d'enveloppe temporelle en faisant appel à une table de codage représentant les enveloppes les plus couramment rencontrées,
- 30 • coder le « pitch » en loi logarithmique sur sa valeur ou la différence entre le « pitch » codé de la trame précédente et celui de la trame

courante ; à noter que le codage différentiel permet d'utiliser moins de bits de codage,

- 5 • coder la variation du « pitch », n'étant pas en présence d'une impulsion, seulement si la valeur calculée dans l'analyse est éloignée d'un certain pourcentage de la variation du « pitch » calculée à partir des « pitches » de la trame précédente et de la trame courante ; de même, la variation du « pitch » n'est pas codée, si la valeur absolue de l'écart entre ces deux variations est inférieure à une valeur maximale,
- 10 • coder le formant différentiel sur 2 bits pour les fréquences basses, et sur 1 bit pour les autres fréquences, le premier formant n'étant pas codé en différentiel. A noter que plus il y a d'échantillons de formant à coder, meilleure est la qualité auditive du codeur à débit fixe, et plus la différence de codage entre deux échantillons
- 15 adjacents est faible.

Le décodage étant l'inverse du codage, le « pitch » fournit tous les harmoniques de la voix ; leurs amplitudes sont celles du formant.

- 20 Quant aux fréquences du signal non voisé, on calcule des fréquences espacées entre elles par une valeur moyenne à laquelle est rajouté un écart aléatoire ; les amplitudes sont celles du formant.

Le processus de synthèse, décrit précédemment, est identique à celui décrit pour un décodeur à débit variable.

25

Afin de permettre la mise en œuvre du procédé selon l'invention, un dispositif sera décrit ci-après, avec référence à la figure 5.

Le dispositif, selon l'invention, comprend essentiellement :

- 30 • un calculateur 71, de type DSP, permettant d'effectuer le traitement numérique des signaux,

- un clavier 72 permettant de sélectionner les menus de traitement de la voix,
- une mémoire morte 73, de type EEPROM, contenant le logiciel de traitement de la voix,
- 5 • une mémoire vive 74, de type flash ou « memory stick », contenant les enregistrements de la voix traitée,
- un afficheur 75, de type LCD, associé au clavier 72, indiquant les différents menus de traitement de la voix,
- un codeur/décodeur 76, de type codec, assurant les liaisons
- 10 entrée/sortie des périphériques audio,
- un micro 77, de type électret,
- un haut-parleur 78,
- une batterie 79,
- une liaison entrée/sortie 80, permettant le transfert des
- 15 enregistrements numériques et les mises à jour du logiciel de traitement de la voix.

Par ailleurs, le dispositif pourra comporter :

- un connecteur téléphonique permettant au dispositif selon
- 20 l'invention de se substituer à un combiné téléphonique,
- un connecteur de téléphonie mobile,
- une sortie casque, permettant l'écoute des enregistrements,
- une sortie chaîne hi fi, permettant la fonction karaoké,
- un connecteur d'alimentation externe.

25

D'une manière plus précise, le dispositif pourra comporter :

- des moyens d'analyse permettant de déterminer des paramètres représentatifs dudit signal sonore, les susdits moyens d'analyse comprenant :
- 30 • des moyens de calcul de l'enveloppe du signal,

- des moyens de calcul du « pitch » et de sa variation,
 - des moyens d'application au signal temporel de la variation inverse du «pitch»,
 - des moyens de transformée rapide de Fourier (TRF) sur le signal
5 prétraité,
 - des moyens d'extraction des composantes fréquentielles et leurs amplitudes dudit signal, à partir du résultat de la transformée rapide de Fourier,
 - des moyens d'élimination éventuelle du bruit ambiant par filtrage
10 sélectif avant codage,
- des moyens de synthèse desdits paramètres représentatifs permettant de reconstituer ledit signal sonore, les susdits moyens de synthèse comprenant :
- des moyens de sommation des sinus dont l'amplitude des
15 composantes fréquentielles varie en fonction de l'enveloppe du signal,
 - des moyens de calcul des phases en fonction de la valeur des fréquences et des valeurs des phases et des fréquences appartenant à
20 la trame précédente,
 - des moyens de superposition du bruit,
 - des moyens d'application de l'enveloppe,
- des moyens de filtrage du bruit et de génération d'effets spéciaux, à partir
25 de l'analyse, sans passer par la synthèse, les susdits moyens de filtrage du bruit et de génération d'effets spéciaux comprenant :
- des moyens de sommation du signal original, du signal original décalé de un « pitch » en valeur positive et du signal original décalé de un « pitch » en valeur négative,
 - des moyens de division de la valeur temporelle du « pitch » par
30 deux,

- des moyens de modification des amplitudes du signal original et des deux signaux décalés,
 - des moyens de multiplication de chaque échantillon de la voix originale par un cosinus variant au rythme de la moitié de la fondamentale (multiplication par deux du nombre de fréquences), ou
5 variant au rythme du tiers de la fondamentale (multiplication par trois du nombre de fréquences),
 - des moyens d'addition ensuite du résultat obtenu à la voix originale,
- 10 - des moyens de génération d'effets spéciaux associés à la synthèse, les susdits moyens de génération d'effets spéciaux comprenant :
- des moyens de multiplication de toutes les fréquences des composantes fréquentielles du signal original, prises individuellement, par un coefficient,
 - 15 • des moyens de régénération des modules des harmoniques à partir de l'enveloppe spectrale dudit signal original.

Avantageusement, le dispositif pourra comporter tous les éléments cités précédemment, en version professionnelle ou semi professionnelle ; certains
20 éléments, tel que l'afficheur, pourront être simplifiés en version de base.

Ainsi, le dispositif selon l'invention, décrit ci-dessus, pourra exploiter le procédé de traitement numérique différencié de la voix et de la musique, de filtrage du bruit et la création d'effets spéciaux.

25

Il permettra notamment de transformer la voix :

- en une autre voix réaliste,
- pour un usage de type karaoké,
- en une autre voix futuriste, étrange, d'accompagnement.

30

Il permettra également :

- de supprimer le bruit ambiant et d'augmenter les capacités d'enregistrement,
- 5 • de transférer les enregistrements sur disque dur d'ordinateur et de les réécouter à vitesse variable,
- de réaliser une fonction « main libre » associée à un radiotéléphone mobile.
- de générer une réponse auditive adaptée aux mal entendants.

Revendications

- 1- Procédé pour le traitement numérique différencié d'un signal sonore, constitué dans l'intervalle d'une trame par la somme de sinus d'amplitude fixe
5 et dont la fréquence est modulée linéairement en fonction du temps, cette somme étant modulée temporellement par une enveloppe, le bruit dudit signal sonore étant rajouté audit signal, préalablement à ladite somme, caractérisé en ce qu'il comprend :
- une étape d'analyse permettant de déterminer des paramètres représentatifs
10 dudit signal sonore, par
- un calcul de l'enveloppe du signal,
 - un calcul de la période du fondamental du signal de voix (« pitch ») et de sa variation,
 - une application au signal temporel de la variation inverse du «pitch»,
 - 15 • une transformée rapide de Fourier (TRF) sur le signal prétraité,
 - une extraction des composantes fréquentielles et leurs amplitudes dudit signal, à partir du résultat de la transformée rapide de Fourier,
 - un calcul du «pitch» et sa validation dans le domaine fréquentiel,
 - une élimination éventuelle du bruit ambiant par filtrage sélectif avant
20 codage.

- 2- Procédé selon la revendication 1,
caractérisé en ce qu'il comprend en outre une étape de synthèse desdits
paramètres représentatifs permettant de reconstituer ledit signal sonore.
25

- 3- Procédé selon les revendications précédentes,
caractérisé en ce qu'il comprend en outre une étape de codage et de décodage
desdits paramètres représentatifs dudit signal sonore.

4- Procédé selon les revendications précédentes, caractérisé en ce qu'il comprend en outre une étape de filtrage du bruit et une étape de génération d'effets spéciaux, à partir de l'analyse, sans passer par la synthèse.

5

5- Procédé selon les revendications précédentes, caractérisé en ce qu'il comprend en outre une étape de génération d'effets spéciaux associés à la synthèse.

10

6- Procédé selon la revendication 2, caractérisé en ce que la susdite étape de synthèse comprend :

- une sommation des sinus dont l'amplitude des composantes fréquentielles varie en fonction de l'enveloppe du signal et dont les fréquences varient linéairement,
- 15 • un calcul des phases en fonction de la valeur des fréquences et des valeurs des phases et des fréquences appartenant à la trame précédente,
- une superposition du bruit,
- une application de l'enveloppe.

20

7- Procédé selon la revendication 4, caractérisé en ce que la susdite étape de filtrage du bruit et la susdite étape de génération d'effets spéciaux, à partir de l'analyse, sans passer par la synthèse, comprennent une somme du signal original, du signal original décalé de un
25 « pitch » en valeur positive et du signal original décalé de un « pitch » en valeur négative.

8- Procédé selon la revendication 7, caractérisé en ce que les susdits signaux décalés sont multipliés par un même
30 coefficient, et le signal original par un second coefficient, la somme dudit

premier coefficient, rajouté à lui-même, et dudit second coefficient est égale à 1, diminué de manière à conserver un niveau équivalent du signal résultant.

9- Procédé selon la revendication 7,

5 caractérisé en ce que la susdite étape de filtrage et la susdite étape de génération d'effets spéciaux, à partir de l'analyse, sans passer par la synthèse, comprennent :

- une division de la valeur temporelle du « pitch » par deux,
 - une modification des amplitudes du signal original et des deux
- 10 signaux décalés.

10- Procédé selon la revendication 7,

caractérisé en ce que la susdite étape de filtrage et la susdite étape de génération d'effets spéciaux, à partir de l'analyse, sans passer par la synthèse,

15 comprennent :

- une multiplication de chaque échantillon de la voix originale par un cosinus variant au rythme de la moitié de la fondamentale (multiplication par deux du nombre de fréquences), ou variant au rythme du tiers de la fondamentale (multiplication par trois du
- 20 nombre de fréquences),
- une addition ensuite du résultat obtenu à la voix originale.

11- Procédé selon la revendication 5,

caractérisé en ce que la susdite étape de génération d'effets spéciaux associés à la synthèse, comprend :

25

- une multiplication de toutes les fréquences des composantes fréquentielles du signal original, prises individuellement, par un coefficient,
 - une régénération des modules des harmoniques à partir de
- 30 l'enveloppe spectrale dudit signal original.

12- Procédé selon la revendication 11,
caractérisé en ce que le susdit coefficient de multiplication des composantes
fréquentielles est :

- 5 • un coefficient fonction du rapport entre le nouveau « pitch » et le
 « pitch » réel,
- un coefficient variant, périodiquement ou aléatoirement, à basse
 fréquence.

13- Dispositif, pour la mise en œuvre du procédé selon la revendication
10 1, de traitement numérique différencié d'un signal sonore, constitué dans
l'intervalle d'une trame par la somme de sinus d'amplitude fixe et dont la
fréquence est modulée linéairement en fonction du temps, cette somme étant
modulée temporellement par une enveloppe, le bruit dudit signal sonore étant
rajouté audit signal, préalablement à ladite somme,

15 caractérisé en ce qu'il comprend :

- des moyens d'analyse permettant de déterminer des paramètres
représentatifs dudit signal sonore, et/ou
- des moyens de synthèse desdits paramètres représentatifs permettant de
reconstituer ledit signal sonore, et/ou
- 20 - des moyens de codage et de décodage desdits paramètres représentatifs
dudit signal sonore, et/ou
- des moyens de filtrage du bruit et de génération d'effets spéciaux, à partir
de l'analyse, sans passer par la synthèse, et/ou
- des moyens de génération d'effets spéciaux associés à la synthèse.

25

14- Dispositif selon la revendication 13,
caractérisé en ce que les susdits moyens d'analyse comprennent :

- des moyens de calcul de l'enveloppe du signal,
- des moyens de calcul du « pitch » et de sa variation,
- 30 • des moyens d'application au signal temporel de la variation inverse
du «pitch»,

- des moyens de transformée rapide de Fourier (TRF) sur le signal prétraité,
- des moyens d'extraction des composantes fréquentielles et leurs amplitudes dudit signal, à partir du résultat de la transformée rapide de Fourier,
- des moyens d'élimination éventuelle du bruit ambiant par filtrage sélectif avant codage.

- 15- Dispositif selon la revendication 13,
- 10 caractérisé en ce que les susdits moyens de synthèse comprennent :
- des moyens de sommation des sinus dont l'amplitude des composantes fréquentielles varie en fonction de l'enveloppe du signal,
 - des moyens de calcul des phases en fonction de la valeur des fréquences et des valeurs des phases et des fréquences appartenant à la trame précédente,
 - des moyens de superposition du bruit,
 - des moyens d'application de l'enveloppe.

- 20 16- Dispositif selon la revendication 13,
- caractérisé en ce que les susdits moyens de filtrage du bruit et de génération d'effets spéciaux, à partir de l'analyse, sans passer par la synthèse, comprennent des moyens de sommation du signal original, du signal original décalé de un « pitch » en valeur positive et du signal original décalé de un
- 25 « pitch » en valeur négative.

- 17- Dispositif selon la revendication 16,
- caractérisé en ce que les susdits signaux décalés sont multipliés par un même coefficient, et le signal original par un second coefficient, la somme dudit
- 30 premier coefficient, rajouté à lui-même, et dudit second coefficient est égale à 1, diminué de manière à conserver un niveau équivalent du signal résultant.

18- Dispositif selon la revendication 13,
caractérisé en ce que les susdits moyens de filtrage et de génération d'effets
spéciaux, à partir de l'analyse, sans passer par la synthèse, comprennent :

- 5
- des moyens de division de la valeur temporelle du « pitch » par deux,
 - des moyens de modification des amplitudes du signal original et des deux signaux décalés.

10 19- Dispositif selon la revendication 13,
caractérisé en ce que les susdits moyens de filtrage et de génération d'effets
spéciaux, à partir de l'analyse, sans passer par la synthèse, comprennent :

- 15
- des moyens de multiplication de chaque échantillon de la voix originale par un cosinus variant au rythme de la moitié de la fondamentale (multiplication par deux du nombre de fréquences), ou variant au rythme du tiers de la fondamentale (multiplication par trois du nombre de fréquences),
 - des moyens d'addition ensuite du résultat obtenu à la voix originale.

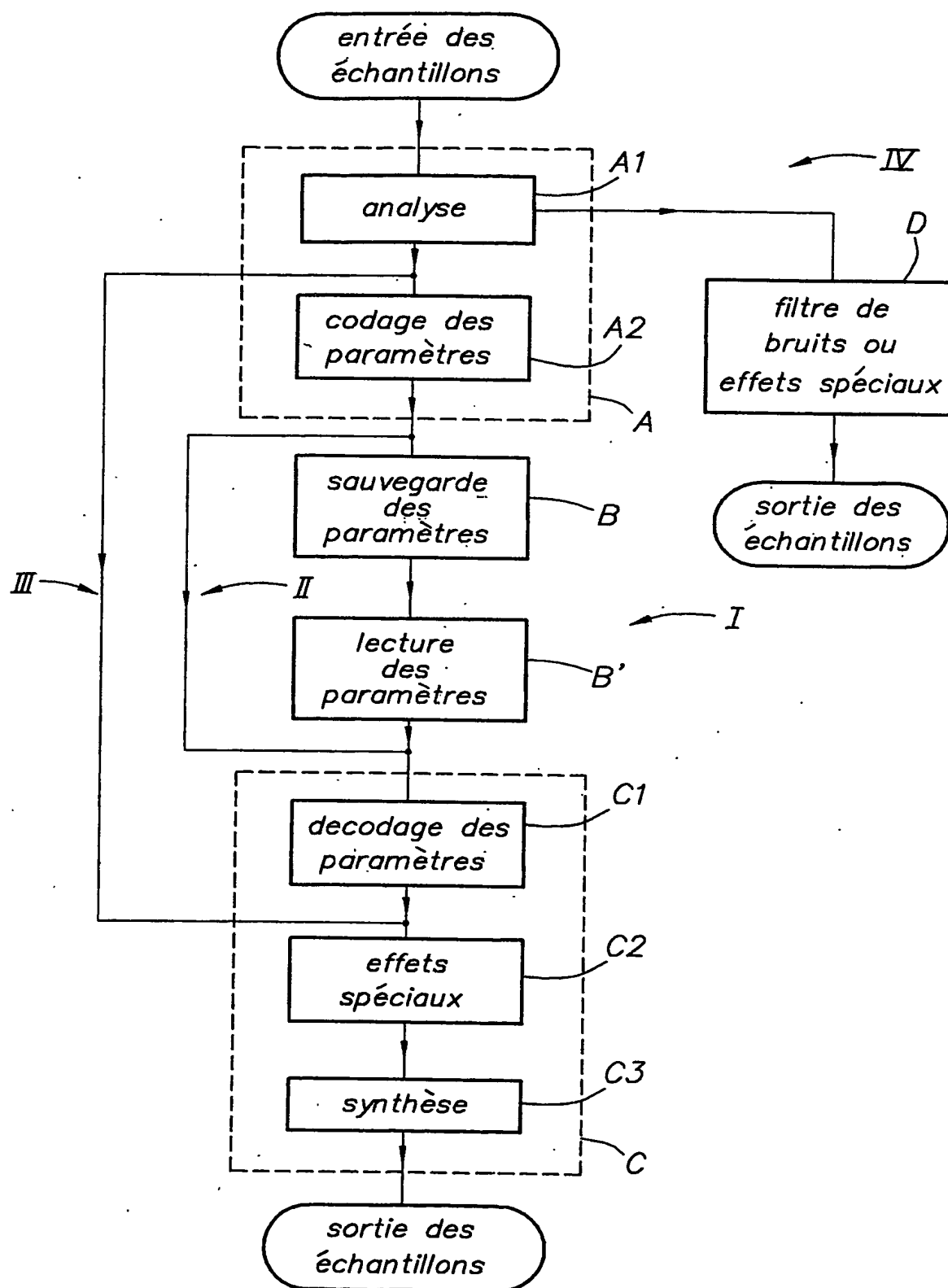
20 20- Dispositif selon la revendication 13,
caractérisé en ce que les susdits moyens de génération d'effets spéciaux
associés à la synthèse, comprennent :

- 25
- des moyens de multiplication de toutes les fréquences des composantes fréquentielles du signal original, prises individuellement, par un coefficient,
 - des moyens de régénération des modules des harmoniques à partir de l'enveloppe spectrale dudit signal original.

21- Dispositif selon la revendication 20,
caractérisé en ce que le susdit coefficient de multiplication des composantes
fréquentielles est :

- un coefficient fonction du rapport entre le nouveau « pitch » et le
5 « pitch » réel,
- un coefficient variant périodiquement, à basse fréquence.

FIG.1



2/5

FIG.2

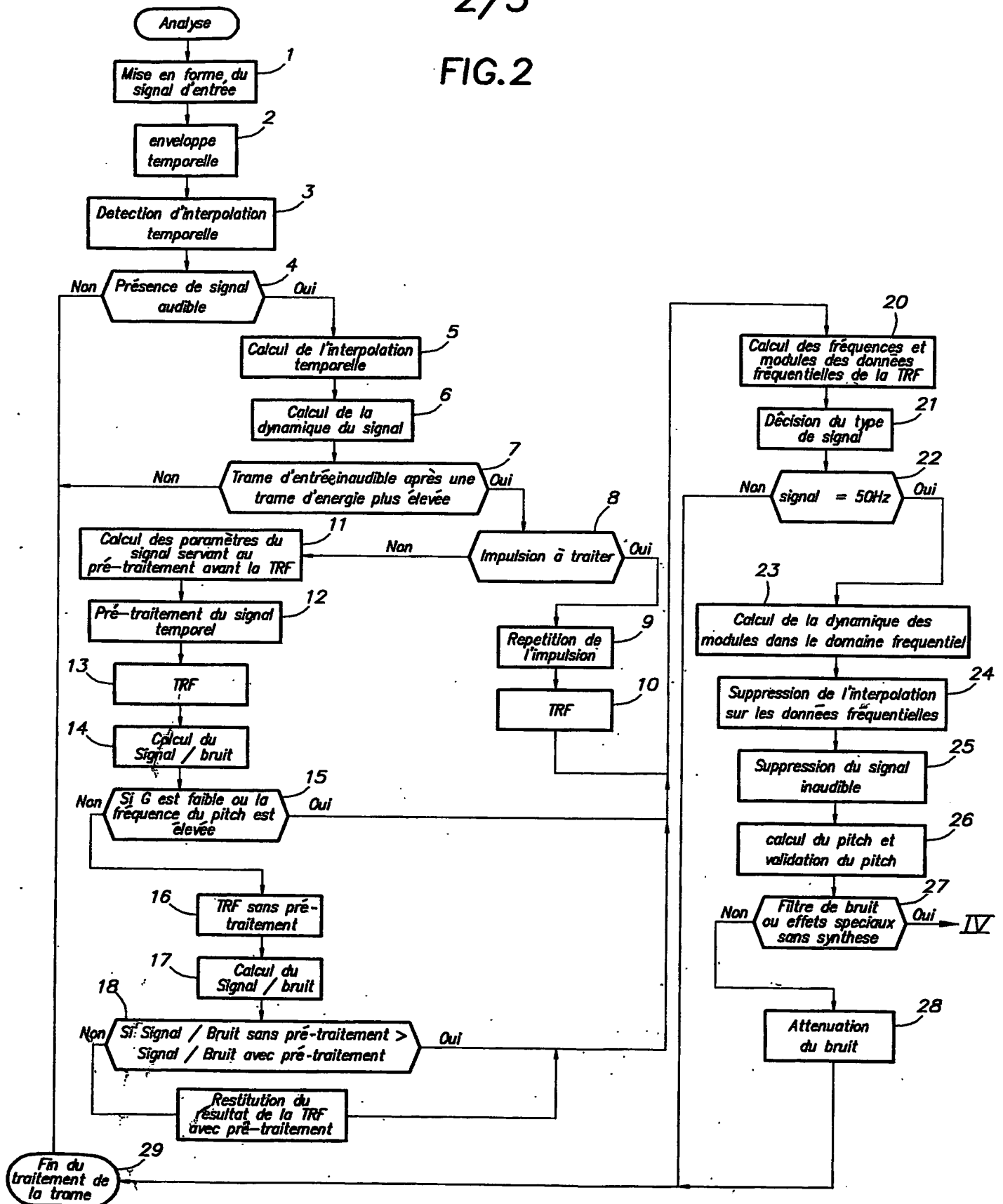


FIG.3

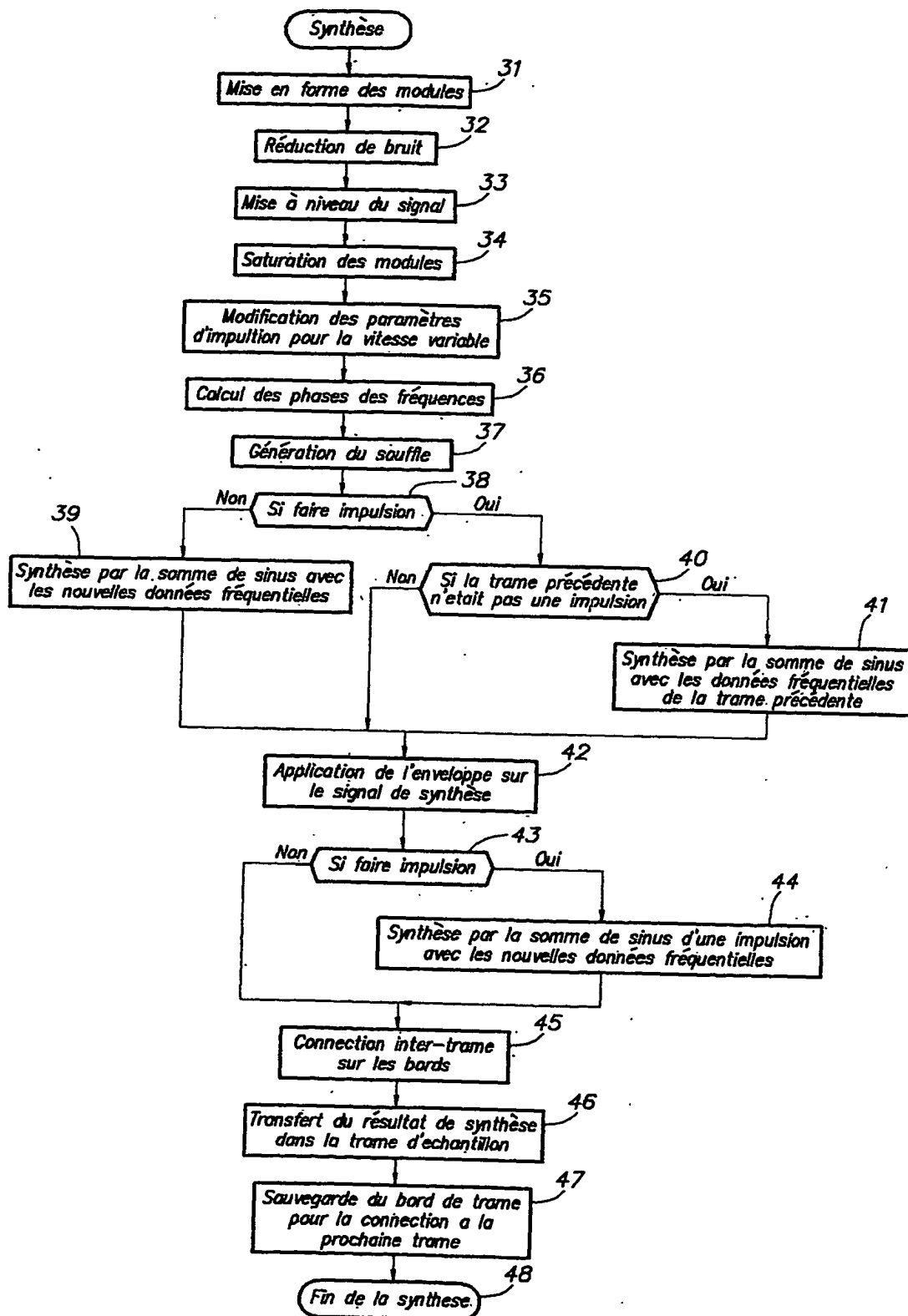


FIG.4

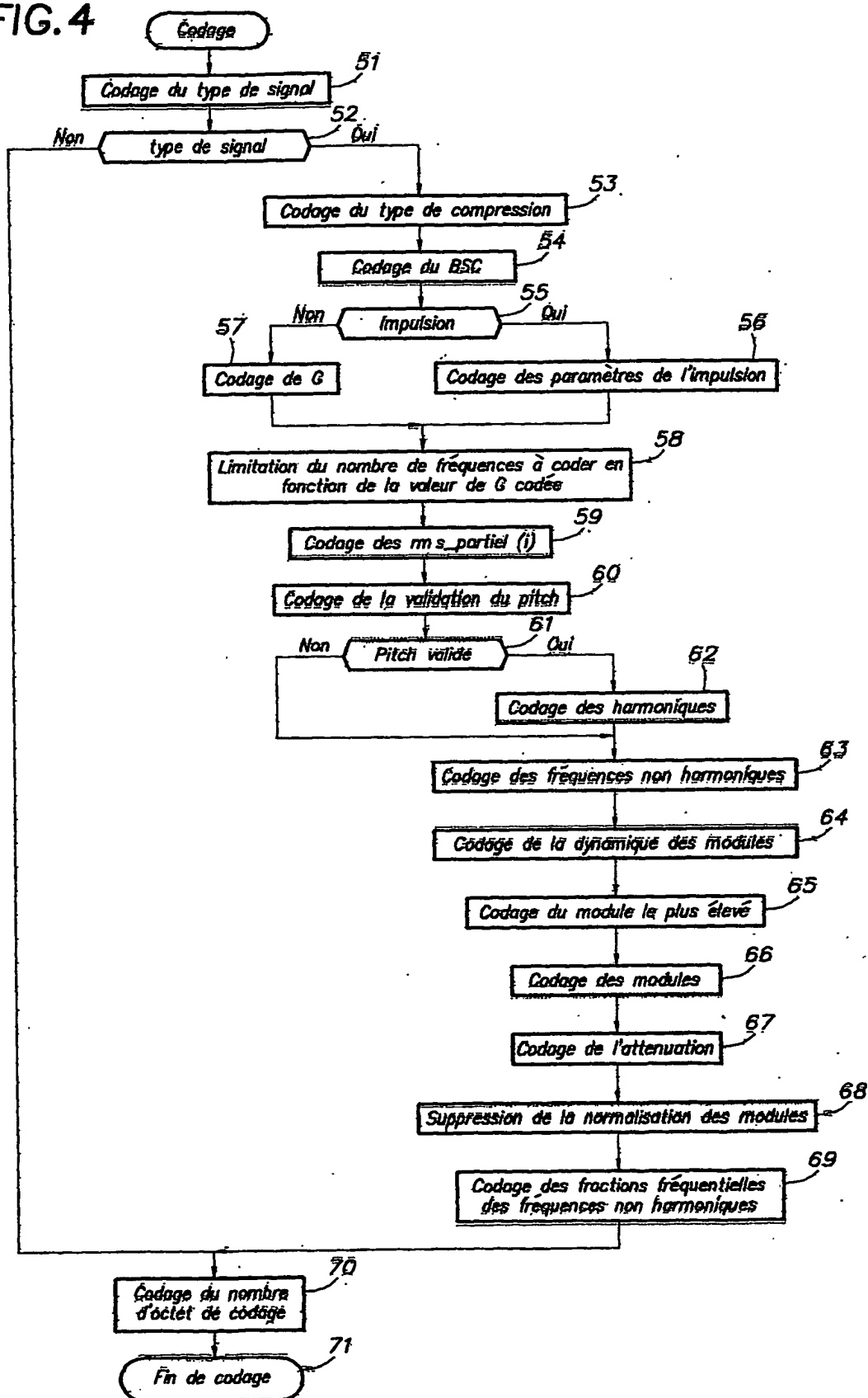
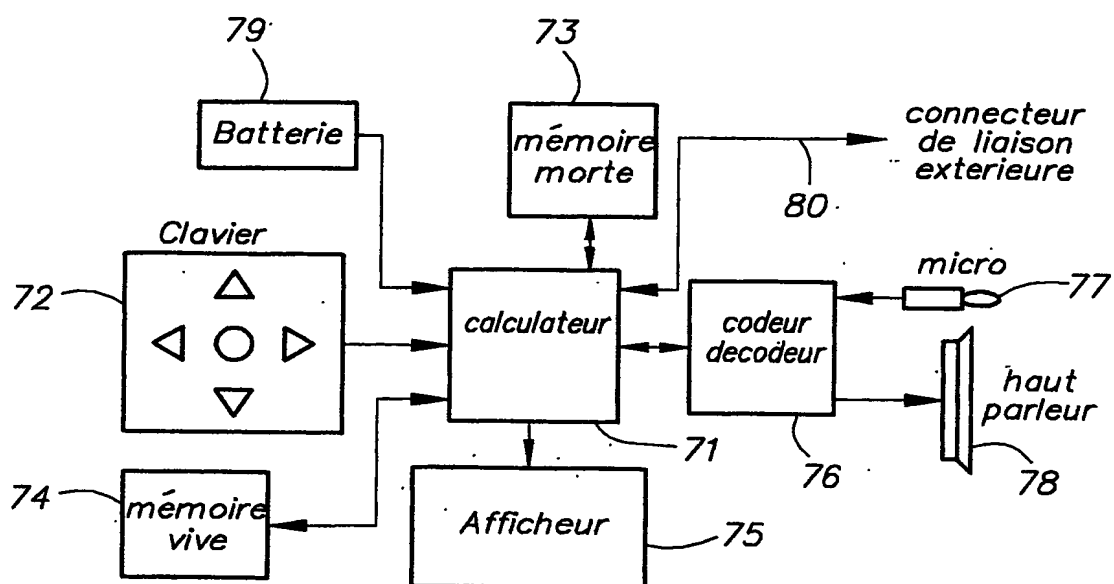


FIG. 5



INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR2004/000184

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L19/02 G10L21/02

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 01/59766 A (COMSAT CORP ;YELDENER SUAT (US)) 16 August 2001 (2001-08-16) abstract; figure 3	1-6, 11, 13-15, 20
Y	US 2002/184009 A1 (HEIKKINEN ARI P) 5 December 2002 (2002-12-05) abstract; figure 7 paragraphs '0004! - '0011! paragraph '0013! ----- -/-	1-6, 13-15

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

7 June 2004

Date of mailing of the international search report

17/06/2004

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Quélavoine, R

INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR2004/000184

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MOULINES E ET AL: "Non-parametric techniques for pitch-scale and time-scale modification of speech" SPEECH COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 16, no. 2, 1 February 1995 (1995-02-01), pages 175-205, XP004024959 ISSN: 0167-6393 abstract * section 2.3.2 Pitch scale modification *	11,20
A	US 5 684 262 A (NAKAMURA ET AL) 4 November 1997 (1997-11-04) abstract	7,8,16, 17

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/FR2004/000184

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
WO 0159766	A	16-08-2001	AU	4147501 A		20-08-2001
			CA	2399706 A1		16-08-2001
			WO	0159766 A1		16-08-2001
<hr/>						
US 2002184009	A1	05-12-2002	EP	1390945 A1		25-02-2004
			WO	02097798 A1		05-12-2002
<hr/>						
US 5684262	A	04-11-1997	JP	3351905 B2		03-12-2002
			JP	8044389 A		16-02-1996
			CN	1131471 A ,B		18-09-1996
			WO	9603740 A1		08-02-1996
<hr/>						

RAPPORT DE RECHERCHE INTERNATIONALE

Demande Internationale No

PCT/FR2004/000184

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
CIB 7 G10L19/02 G10L21/02

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)

CIB 7 G10L

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)

EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
Y	WO 01/59766 A (COMSAT CORP ; YELDENER SUAT (US)) 16 août 2001 (2001-08-16) abrégé; figure 3	1-6, 11, 13-15, 20
Y	US 2002/184009 A1 (HEIKKINEN ARI P) 5 décembre 2002 (2002-12-05) abrégé; figure 7 alinéas '0004! - '0011! alinéa '0013!	1-6, 13-15

☒ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

- 'A' document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- 'E' document antérieur, mais publié à la date de dépôt international ou après cette date
- 'L' document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- 'O' document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- 'P' document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

- 'T' document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention
- 'X' document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément
- 'Y' document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier
- '&' document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

7 juin 2004

Date d'expédition du présent rapport de recherche internationale

17/06/2004

Nom et adresse postale de l'administration chargée de la recherche internationale
Office Européen des Brevets, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Quélavoine, R

RAPPORT DE RECHERCHE INTERNATIONALE

Demande Internationale No

PCT/FR2004/000184

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie °	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
Y	<p>MOULINES E ET AL: "Non-parametric techniques for pitch-scale and time-scale modification of speech"</p> <p>SPEECH COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL,</p> <p>vol. 16, no. 2,</p> <p>1 février 1995 (1995-02-01), pages 175-205, XP004024959</p> <p>ISSN: 0167-6393</p> <p>abrégé</p> <p>* section 2.3.2 Pitch scale modification *</p> <p>-----</p>	11,20
A	<p>US 5 684 262 A (NAKAMURA ET AL)</p> <p>4 novembre 1997 (1997-11-04)</p> <p>abrégé</p> <p>-----</p>	7,8,16,17

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande internationale No

PCT/FR2004/000184

Document brevet cité au rapport de recherche		Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 0159766	A	16-08-2001	AU 4147501 A	20-08-2001
			CA 2399706 A1	16-08-2001
			WO 0159766 A1	16-08-2001
US 2002184009	A1	05-12-2002	EP 1390945 A1	25-02-2004
			WO 02097798 A1	05-12-2002
US 5684262	A	04-11-1997	JP 3351905 B2	03-12-2002
			JP 8044389 A	16-02-1996
			CN 1131471 A , B	18-09-1996
			WO 9603740 A1	08-02-1996